PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : C07K	A2	(11) International Publication Number: WO 99/01468
		(43) International Publication Date: 14 January 1999 (14.01.99)

(21) International Application Number: PCT/US98/13793

(22) International Filing Date: 1 July 1998 (01.07.98)

(30) Priority Data:

60/051,407	1 July 1997 (01.07.97)	US
08/897,798	21 July 1997 (21.07.97)	US

(71) Applicants: ONTOGENY, INC. [US/US]; 45 Moulton Street, Cambridge, MA 02138 (US). INGHAM, Philip, W. [–/GB]; 88 Stumperlowe Hall Road, S10 3QT (GB). VAN DEN HEUVEL, Marcel [–/GB]; Squirrel Cottage, **, Oxford OX3 GUK (GB).

(74) Agents: VINCENT, Matthew, P. et al.; Foley, Hoag & Eliot LLP, One Post Office Square, Boston, MA 02109 (US).

(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, GW, HR, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

Published

Without international search report and to be republished upon receipt of that report.

(54) Title: VERTEBRATE SMOOTHENED GENE, GENE PRODUCTS, AND USES RELATED THERETO

(57) Abstract

The present invention concerns the discovery of a new family of serpentine receptor proteins, referred to herein as "*smoothened*" proteins. *smoothened* is demonstrated to be involved in *hedgehog* signal transduction, and play an important role in *hedgehog*–mediated induction of tissue.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	**ES**	Spain	**LS**	Lesotho	**SI**	Slovenia
AM	Armenia	**FI**	Finland	**LT**	Lithuania	**SK**	Slovakia
AT	Austria	**FR**	France	**LU**	Luxembourg	**SN**	Senegal
AU	Australia	**GA**	Gabon	**LV**	Latvia	**SZ**	Swaziland
AZ	Azerbaijan	**GB**	United Kingdom	**MC**	Monaco	**TD**	Chad
BA	Bosnia and Herzegovina	**GE**	Georgia	**MD**	Republic of Moldova	**TG**	Togo
BB	Barbados	**GH**	Ghana	**MG**	Madagascar	**TJ**	Tajikistan
BE	Belgium	**GN**	Guinea	**MK**	The former Yugoslav Republic of Macedonia	**TM**	Turkmenistan
BF	Burkina Faso	**GR**	Greece			**TR**	Turkey
BG	Bulgaria	**HU**	Hungary	**ML**	Mali	**TT**	Trinidad and Tobago
BJ	Benin	**IE**	Ireland	**MN**	Mongolia	**UA**	Ukraine
BR	Brazil	**IL**	Israel	**MR**	Mauritania	**UG**	Uganda
BY	Belarus	**IS**	Iceland	**MW**	Malawi	**US**	United States of America
CA	Canada	**IT**	Italy	**MX**	Mexico	**UZ**	Uzbekistan
CF	Central African Republic	**JP**	Japan	**NE**	Niger	**VN**	Viet Nam
CG	Congo	**KE**	Kenya	**NL**	Netherlands	**YU**	Yugoslavia
CH	Switzerland	**KG**	Kyrgyzstan	**NO**	Norway	**ZW**	Zimbabwe
CI	Côte d'Ivoire	**KP**	Democratic People's Republic of Korea	**NZ**	New Zealand		
CM	Cameroon			**PL**	Poland		
CN	China	**KR**	Republic of Korea	**PT**	Portugal		
CU	Cuba	**KZ**	Kazakstan	**RO**	Romania		
CZ	Czech Republic	**LC**	Saint Lucia	**RU**	Russian Federation		
DE	Germany	**LI**	Liechtenstein	**SD**	Sudan		
DK	Denmark	**LK**	Sri Lanka	**SE**	Sweden		
EE	Estonia	**LR**	Liberia	**SG**	Singapore		

Vertebrate Smoothened Gene, Gene Products, and Uses Related Thereto

Background of the Invention

Pattern formation is the activity by which embryonic cells form ordered spatial arrangements of differentiated tissues. The physical complexity of higher organisms arises during embryogenesis through the interplay of cell-intrinsic lineage and cell-extrinsic signaling. Inductive interactions are essential to embryonic patterning in vertebrate development from the earliest establishment of the body plan, to the patterning of the organ systems, to the generation of diverse cell types during tissue differentiation (Davidson, E., (1990) *Development* 108: 365-389; Gurdon, J. B., (1992) *Cell* 68: 185-199; Jessell, T. M. et al., (1992) *Cell* 68: 257-270). The effects of developmental cell interactions are varied. Typically, responding cells are diverted from one route of cell differentiation to another by inducing cells that differ from both the uninduced and induced states of the responding cells (inductions). Sometimes cells induce their neighbors to differentiate like themselves (homoiogenetic induction); in other cases a cell inhibits its neighbors from differentiating like itself. Cell interactions in early development may be sequential, such that an initial induction between two cell types leads to a progressive amplification of diversity. Moreover, inductive interactions occur not only in embryos, but in adult cells as well, and can act to establish and maintain morphogenetic patterns as well as induce differentiation (J.B. Gurdon (1992) *Cell* 68:185-199).

The origin of the nervous system in all vertebrates can be traced to the end of gastrulation. At this time, the ectoderm in the dorsal side of the embryo changes its fate from epidermal to neural. The newly formed neuroectoderm thickens to form a flattened structure called the neural plate which is characterized, in some vertebrates, by a central groove (neural groove) and thickened lateral edges (neural folds). At its early stages of differentiation, the neural plate already exhibits signs of regional differentiation along its anterior posterior (A-P) and mediolateral axis (M-L). The neural folds eventually fuse at the dorsal midline to form the neural tube which will differentiate into brain at its anterior end and spinal cord at its posterior end. Closure of the neural tube creates dorsal/ventral differences by virtue of previous mediolateral differentiation. Thus, at the end of neurulation, the neural tube has a clear anterior-posterior (A-P), dorsal ventral (D-V) and mediolateral (M-L) polarities (see, for example, *Principles in Neural Science (3rd)*, eds. Kandel, Schwartz and Jessell, Elsevier Science Publishing Company: NY, 1991; and *Developmental Biology (3rd)*, ed. S.F. Gilbert, Sinauer Associates: Sunderland MA, 1991). Inductive interactions that define the fate of cells within the neural tube establish the initial pattern of the embryonic vertebrate nervous system. In the spinal cord, the identify of cell types is controlled, in part, by signals from two midline cell groups, the notochord and floor

plate, that induce neural plate cells to differentiate into floor plate, motor neurons, and other ventral neuronal types (van Straaten et al. (1988) *Anat. Embryol.* 177:317-324; Placzek et al. (1993) *Development* 117:205-218; Yamada et al. (1991) *Cell* 64:035-647; and Hatta et al. (1991) *Nature* 350:339-341). In addition, signals from the floor plate are responsible for the orientation and direction of commissural neuron outgrowth (Placzek, M. et al., (1990) *Development* 110: 19-30). Besides patterning the neural tube, the notochord and floorplate are also responsible for producing signals which control the patterning of the somites by inhibiting differentiation of dorsal somite derivatives in the ventral regions (Brand-Saberi, B. et al., (1993) *Anat. Embryol.* 188: 239-245; Porquie, O. et al., (1993) *Proc. Natl. Acad. Sci. USA* 90: 5242-5246).

Another important signaling center exists in the posterior mesenchyme of developing limb buds, called the Zone of Polarizing Activity, or "ZPA". When tissue from the posterior region of the limb bud is grafted to the anterior border of a second limb bud, the resultant limb will develop with additional digits in a mirror-image sequence along the anteroposterior axis (Saunders and Gasseling, (1968) *Epithelial-Mesenchymal Interaction*, pp. 78-97). This finding has led to the model that the ZPA is responsible for normal anteroposterior patterning in the limb. The ZPA has been hypothesized to function by releasing a signal, termed a "morphogen", which forms a gradient across the early embryonic bud. According to this model, the fate of cells at different distances from the ZPA is determined by the local concentration of the morphogen, with specific thresholds of the morphogen inducing successive structures (Wolpert, (1969) *Theor. Biol.* 25:1-47). This is supported by the finding that the extent of digit duplication is proportional to the number of implanted ZPA cells (Tickle, (1981) *Nature* 254:199-202).

Although the existence of inductive signals in the ZPA has been known for years, the molecular identities of these signals are only now beginning to be elucidated. An important step forward has been the discovery that the secreted protein *Sonic hedgehog* (*Shh*) is produced in several tissues with organizing properties, including notochord, floor plate and ZPA (Echelard et al. (1993), *Cell* 75: 1417-1430; Bitgood, M.J. and A.P. McMahon (1995) *Dev. Biol.* 172:126-38). Misexpressing *Shh* mimics the inductive effects on ectopic notochord in the neural tube and somites (Echelard et al. (1993) *supra*) and also mimics ZPA function in the limb bud (Riddle et al. (1993) *Cell* 75:1401-16; Chang et al. (1994) *Development* 120: 3339-53).

The vertebrate hedgehog family of inducing molecules comprises three homologs named Sonic, Indian and Desert *hedgehog* (Riddle et al. (1993) *supra*). Desert *hedgehog* (*Dhh*) is expressed principally in the testes, both in mouse embryonic development and in the adult rodent and human; Indian *hedgehog* (*Ihh*) is involved in bone development during embryogenesis and in bone formation in the adult; and, *Shh*, which as described above, is

primarily involved in morphogenic and neuroinductive activities. Given the critical inductive roles of *hedgehog* polypeptides in the development and maintenance of vertebrate organs, the identification of hedgehog interacting proteins is of paramount significance in both clinical and research contexts.

Summary of the Invention

The present invention relates to the discovery of a new class of *serpentine receptors*, referred to herein as *smoothened* proteins. The *smoothened* proteins of the present invention include polypeptides which affect the transmission of signals by the products of the *hedgehog* gene family. *Hedgehog* family members are known for their broad involvement in the formation and maintenance of ordered spatial arrangements of differentiated tissues in vertebrates, both adult and embryonic, and can be used to generate and/or maintain an array of different vertebrate tissue both *in vitro* and *in vivo*.

In general, the invention features isolated *smoothened* polypeptides, preferably substantially pure preparations of the subject *smoothened* polypeptides, such as liposomal preparations. The invention also provides recombinantly produced *smoothened* polypeptides.

In one embodiment, the polypeptide is identical with or similar to a *smoothened* polypeptide represented in SEQ ID No: 5, SEQ ID No: 6, SEQ ID No: 7 and SEQ ID No: 8. Related members of the *smoothened* family are also contemplated, for instance, a *smoothened* polypeptide preferably has an amino acid sequence at least 65%, 70%, 75% or 80% identical or similar to a polypeptide represented by SEQ ID No: 5, SEQ ID No: 6, SEQ ID No: 7 and SEQ ID No: 8 though polypeptides with higher sequence homologies of, for example, 85, 90% and 95% or are also contemplated. In a preferred embodiment, the *smoothened* polypeptide is encoded by a nucleic acid which hybridizes under stringent conditions with a nucleic acid sequence represented in any one or more of SEQ ID Nos: 1-4 and 9. Homologs of the subject *smoothened* proteins also include versions of the protein which are resistant to post-translation modification, as for example, due to mutations which alter modification sites (such as tyrosine, threonine, serine or aspargine residues), or which prevent glycosylation of the protein, or which prevent interaction of the protein with a *smoothened* ligand, e.g. a *hedgehog* polypeptide.

The *smoothened* polypeptide can comprise a full length protein, such as represented in SEQ ID No: 5, SEQ ID No: 6 or SEQ ID No: 7, or it can comprise a fragment corresponding to one or more particular motifs/domains, or to arbitrary sizes, e.g., at least 5, 10, 25, 50, 100, 150 or 200 amino acids in length. In a preferred embodiment, the *smoothened* polypeptide includes at least a portion of a *smoothened* protein corresponding

to Met 13 - Ser 1035 of SEQ ID No. 5. In other preferred embodiments, the *smoothened* polypeptide includes a sufficient portion of the protein to be able to specifically bind to *patched*. Truncated forms of the protein include, but are not limited to, soluble extracellular and/or intracellular fragments, e.g., which bind to ligand or signal transduction proteins, respectively.

The subject proteins can also be provided as chimeric molecules, such as in the form of fusion proteins. For instance, the *smoothened* protein can be provided as a recombinant fusion protein which includes a second polypeptide portion, e.g., a second polypeptide having an amino acid sequence unrelated (heterologous) to the *smoothened* polypeptide.

In yet another embodiment, the invention features nucleic acids encoding *smoothened* polypeptides, which have the ability to modulate, e.g., either mimic or antagonize, at least a portion of the activity of a wild-type *smoothened* polypeptide. Exemplary *smoothened*-encoding nucleic acid sequences are represented by SEQ ID No: 1, SEQ ID No: 2, SEQ ID No: 3 or SEQ ID No: 4.

In another embodiment, the nucleic acids of the present invention include coding sequences which hybridize under medium or high stringency conditions with all or a portion of the coding sequences designated in one or more of SEQ ID Nos: 1-4. The coding sequences of the nucleic acids can comprise sequences which are identical to coding sequences represented in SEQ ID Nos: 1-4, or it can merely be homologous to those sequences. In preferred embodiments, the nucleic acids encode polypeptides which specifically modulate, by acting as either agonists or antagonists, one or more of the bioactivities of wild-type *smoothened* polypeptides.

Furthermore, in certain preferred embodiments, the subject *smoothened* nucleic acids will include a transcriptional regulatory sequence, e.g. at least one of a transcriptional promoter or transcriptional enhancer sequence, which regulatory sequence is operably linked to the *smoothened* gene sequences. Such regulatory sequences can be used in to render the *smoothened* gene sequences suitable for use as an expression vector. This invention also contemplates the cells transfected with said expression vector whether prokaryotic or eukaryotic and a method for producing *smoothened* proteins by employing said expression vectors.

In yet another embodiment, the nucleic acid hybridizes under stringent conditions to nucleic acid probes corresponding to at least 12 consecutive nucleotides of either sense or antisense sequences of SEQ ID No: 1, SEQ ID No: 2, SEQ ID No: 3 and SEQ ID No: 4; though preferably to at least 25 consecutive nucleotides; and more preferably to at least 40, 50 or 75 consecutive nucleotides of either sense or antisense sequence of SEQ ID No: 1, SEQ ID No: 2, SEQ ID No: 3 and SEQ ID No: 4.

Yet another aspect of the present invention concerns an immunogen comprising a *smoothened* polypeptide in an immunogenic preparation, the immunogen being capable of eliciting an immune response specific for a *smoothened* polypeptide; e.g. a humoral response, e.g. an antibody response; e.g. a cellular response. In preferred embodiments, the immunogen comprising an antigenic determinant, e.g. a unique determinant, from a protein represented by one of SEQ ID No: 5, SEQ ID No: 6. SEQ ID No: 7 and/or SEQ ID No: 8.

A still further aspect of the present invention features antibodies and antibody preparations specifically reactive with an epitope of the *smoothened* immunogen.

The invention also features transgenic non-human animals, e.g. mice, rats, rabbits, chickens, frogs or pigs, having a transgene, e.g., animals which include (and preferably express) a heterologous form of a *smoothened* gene described herein, or which misexpress an endogenous *smoothened* gene, e.g., an animal in which expression of one or more of the subject *smoothened* proteins is disrupted. Such a transgenic animal can serve as an animal model for studying cellular and tissue disorders comprising mutated or mis-expressed *smoothened* alleles or for use in drug screening.

The invention also provides a probe/primer comprising a substantially purified oligonucleotide, wherein the oligonucleotide comprises a region of nucleotide sequence which hybridizes under stringent conditions to at least 12 consecutive nucleotides of sense or antisense sequences of any one or more of SEQ ID Nos: 1-4 and 9-14, or naturally occurring mutants thereof. In preferred embodiments, the probe/primer further includes a label group attached thereto and able to be detected. The label group can be selected, e.g., from a group consisting of radioisotopes, fluorescent compounds, enzymes, and enzyme co-factors. Probes of the invention can be used as a part of a diagnostic test kit for identifying dysfunctions associated with mis-expression of a *smoothened* protein, such as for detecting in a sample of cells isolated from a patient, a level of a nucleic acid encoding a *smoothened* protein; e.g. measuring a *smoothened* mRNA level in a cell, or determining whether a genomic *smoothened* gene has been mutated or deleted. These so-called "probes/primers" of the invention can also be used as a part of "antisense" therapy which refers to administration or *in situ* generation of oligonucleotide probes or their derivatives which specifically hybridize (e.g. bind) under cellular conditions, with the cellular mRNA and/or genomic DNA encoding one or more of the subject *smoothened* proteins so as to inhibit expression of that protein, e.g. by inhibiting transcription and/or translation. Preferably, the oligonucleotide is at least 12 nucleotides in length, though primers of 25, 40, 50, or 75 nucleotides in length are also contemplated.

In yet another aspect, the invention provides an assay for screening test compounds for inhibitors, or alternatively, potentiators, of an interaction between a *patched* protein and a *smoothened* polypeptide receptor. In preferred embodiments, the step of detecting

interaction of a target molecule, such as *patched*, and *smoothened* polypeptides is a competitive binding assay. In other preferred embodiments, the step of detecting interaction of the target molecule and *smoothened* polypeptides involves detecting, in a cell-based assay, change(s) in the level of an intracellular second messenger responsive to signaling mediated by the *smoothened* polypeptide. In still another preferred embodiment, the ability to modulate the bioactivity of *smoothened* comprises detecting, in a cell-based assay, change(s) in the level of expression of a gene controlled by a transcriptional regulatory sequence responsive to signaling by the *smoothened* polypeptide.

In preferred embodiments, the steps of the assay are repeated for a variegated library of at least 100 different test compounds, more preferably at least 10^3, 10^4 or 10^5 different test compounds. The test compound can be, e.g., a peptide, a nucleic acid, a carbohydrate, a small organic molecule, or natural product extract (or fraction thereof).

Yet another aspect of the present invention concerns a method for modulating one or more of growth, differentiation, or survival of a cell by modulating *smoothened* bioactivity, e.g., by potentiating or disrupting certain protein-protein interactions. In general, whether carried out *in vivo*, *in vitro*, or *in situ*, the method comprises treating the cell with an effective amount of a *smoothened* therapeutic so as to alter, relative to the cell in the absence of treatment, at least one of (i) rate of growth, (ii) differentiation, or (iii) survival of the cell. Accordingly, the method can be carried out with *smoothened* therapeutics such as peptide and peptidomimetics or other molecules identified in the above-referenced drug screens which agonize or antagonize the effects of signaling from a *smoothened* protein. Other *smoothened* therapeutics include antisense constructs for inhibiting expression of *smoothened* proteins, and dominant negative mutants of *smoothened* proteins which competitively inhibit ligand interactions upstream and signal transduction downstream of the wild-type *smoothened* protein.

In one embodiment, the subject method of modulating *smoothened* bioactivity can be used in the treatment of testicular cells, so as to modulate spermatogenesis. In another embodiment, the subject method is used to modulate osteogenesis, comprising the treatment of osteogenic cells with an agent that modulates *smoothened* bioactivity. Likewise, where the treated cell is a chondrogenic cell, the present method is used to modulate chondrogenesis. In still, another embodiment, the subject method can be used to modulate the differentiation of a neuronal cell, to maintain a neuronal cell in a differentiated state, and/or to enhance the survival of a neuronal cell, e.g., to prevent apoptosis or other forms of cell death. For instance the present method can be used to affect the differentiation of neuronal cells such as motor neurons, cholinergic neurons, dopaminergic neurons, serotonergic neurons, and peptidergic neurons.

Another aspect of the present invention provides a method of determining if a subject, e.g. an animal patient, is at risk for a disorder characterized by unwanted cell proliferation or aberrant control of differentiation or apoptosis. The method includes detecting, in a tissue of the subject, the presence or absence of a genetic lesion characterized by at least one of (i) a mutation of a gene encoding a *smoothened* protein; or (ii) the mis-expression of a *smoothened* gene. In preferred embodiments, detecting the genetic lesion includes ascertaining the existence of at least one of: a deletion of one or more nucleotides from a *smoothened* gene; an addition of one or more nucleotides to the gene, a substitution of one or more nucleotides of the gene, a gross chromosomal rearrangement of the gene; an alteration in the level of a messenger RNA transcript of the gene; the presence of a non-wild type splicing pattern of a messenger RNA transcript of the gene; a non-wild type level of the protein; and/or an aberrant level of soluble *smoothened* protein.

For example, detecting the genetic lesion can include (i) providing a probe/primer including an oligonucleotide containing a region of nucleotide sequence which hybridizes to a sense or antisense sequence of a *smoothened* gene or naturally occurring mutants thereof, or 5' or 3' flanking sequences naturally associated with the *smoothened* gene; (ii) exposing the probe/primer to nucleic acid of the tissue; and (iii) detecting, by hybridization of the probe/primer to the nucleic acid, the presence or absence of the genetic lesion; e.g. wherein detecting the lesion comprises utilizing the probe/primer to determine the nucleotide sequence of the *smoothened* gene and, optionally, of the flanking nucleic acid sequences. For instance, the probe/primer can be employed in a polymerase chain reaction (PCR) or in a ligation chain reaction (LCR). In alternate embodiments, the level of a *smoothened* protein is detected in an immunoassay using an antibody which is specifically immunoreactive with the *smoothened* protein.

The practice of the present invention will employ, unless otherwise indicated, conventional techniques of cell biology, cell culture, molecular biology, transgenic biology, microbiology, recombinant DNA, and immunology, which are within the skill of the art. Such techniques are explained fully in the literature. See, for example, Molecular Cloning A Laboratory Manual, 2nd Ed., ed. by Sambrook, Fritsch and Maniatis (Cold Spring Harbor Laboratory Press: 1989); DNA Cloning, Volumes I and II (D. N. Glover ed., 1985); Oligonucleotide Synthesis (M. J. Gait ed., 1984); Mullis et al. U.S. Patent No: 4,683,195; Nucleic Acid Hybridization (B. D. Hames & S. J. Higgins eds. 1984); Transcription And Translation (B. D. Hames & S. J. Higgins eds. 1984); Culture Of Animal Cells (R. I. Freshney, Alan R. Liss, Inc., 1987); Immobilized Cells And Enzymes (IRL Press, 1986); B. Perbal, A Practical Guide To Molecular Cloning (1984); the treatise, Methods In Enzymology (Academic Press, Inc., N.Y.); Gene Transfer Vectors For Mammalian Cells (J. H. Miller and M. P. Calos eds., 1987, Cold Spring Harbor Laboratory); Methods In Enzymology, Vols. 154 and 155 (Wu et al. eds.),

Immunochemical Methods In Cell And Molecular Biology (Mayer and Walker, eds., Academic Press, London, 1987); Handbook Of Experimental Immunology, Volumes I-IV (D. M. Weir and C. C. Blackwell, eds., 1986); Manipulating the Mouse Embryo, (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1986).

Other features and advantages of the invention will be apparent from the following detailed description, and from the claims.

Brief Description of the Drawings

Figure 1 (panels A-H) are phase-contrast images of the ventral cuticle of pharate first instar larvae; anterior is to the top. *a*, Wild type. *b*, smo^{1LX43}, raised at 25° C. Naked cuticle is deleted sporadically between adjacent denticle belts. *c*, smo^{1LX43}, raised at 18° C. Naked cuticle is replaced by denticle belts with reversed polarity. *d*, *smo-* embryo derived from smo^{D16} germline clone females. All naked cuticle and polarity is lost; this phenotype is typical of all *smo* alleles (three) analyzed in germline clone analysis. *e*, hh^{13C} (severe phenotype [Mohler, J. (1988) *Genetics* 131:643-653]); the phenotype is virtually identical to that of a *smo-* embryo in *d*. *f*, wg^{CX4} (null allele [van den Huevel, M. et al. (1993) *supra*]); note the complete loss of naked cuticle but vestiges of polarity, in contrast, to in *smo-* and *hh* null mutants remain visible. *g*, *smo-; h*GAL4-UAS*wg*; naked cuticle is partially restored in alternate segments. *h*, *smo*-GAL4-AUS*hh*; phenotype is indistinguishable from that of *smo-*.

METHODS: All *smo* alleles have been described (Nüsslein-Volhard, C. et al. (1984) *supra*) except smo^{D16}. This allele and two others (smo^{F5} and smo^{F11}) were isolated in an F_2 lethal screen following -ray mutagenesis. Isogenized *cn bw sp* males were irradiated with 4,000 Rad emitted by a ^{60}Co source and mated to wild-type females. Individual F_1 males (12,000) were test-crossed to smo^{1LX43} *cn bw sp/CyO* females and the F_2 progeny screened for the absence of white-eyed flies. Germline clone females were generated by the dominant female sterile technique either by irradiating larvae with -rays or using flippase-induced mitotic recombination (Chou, T.B. & Perrimon, N. (1992) *Genetics* 131:643-653). Ectopic expression of *wg* and *hh* was induced by the GAL4-UAS system using the *hairy* GAL4 enhancer trap, which expresses GAL4 in every other segment (Capdevilla, J. & Guerrero, I.(1994) *EMBO J.* 13:4459-4468). *h*GAL4 UASwg^{ts}, UAS*hh*, and *h*GAL4 chromosomes have been described.

Figure 2 (panels A-H) show the pattern of expression of *wg*, *hh*, and En in wild-type and mutant embryos. *a*, Expression of *wg* in wild-type stage 10. *b*, Expression of *wg*

(blue/black) and *hh* (red) in *h*GAL4-UAS*hh* embryo. The ectopic expression of *hh* induces the broadening of *wg* in all segments because of the overlap of endogenous *wg* and ectopic *hh*. *c*, *wg* and *hh* in *smo*-; *h*GAL4 UAS*hh*; expression of *eg* in the segmented germband is completely lost as *smo*- embryos (see *d*). *d*, *wg* in *smo*- stage 10; all *wg* expression disappears from the segmented germband. *e*, *h*GAL4 UAS*wg* stage 10 embryo; ectopic and endogenous (arrowheads) *wg* stripes are seen to overlap, with the *h* driven *wg* expression spanning every other segment. *f*, *smo*-; *h*GAL4 UAS*wg* stage 10 embryo; ectopic but no endogenous *wg* expression is detected. *g*, Stage 10 *smo*-; En is lost from the ectoderm in all thoracic and abdominal segments. *h*, Stage 11 *h*GAL4 UAS*wg* embryo, showing slightly broadened stripes of En (compared to wild-type embryos). *i*, *smo*-; *h*GAL4 UAS*wg* stage 10. En expression is maintained in alternate segments (arrowheads), corresponding to the segments where *wg* expression is driven by *h*GAL4. In the intervening stripes, some En is rescued, presumably due to the paracrine action of the misexpressed *wg*.

METHODS: Embryos were collected for one hour and aged for the appropriate times. Fixations, *in situ* hybridizations and antibody stainings (for En, using monoclonal 4D9) have been described (Ingham, P.W. et al. (1991) *supra*; van den Huevel, M. et al. (1993) *supra*). Double-labeling with an *hh* flourescein-labeled and a *wg* digoxigenin probe was performed with two sequential alkaline phosphate substrates, BCIP/NBT and Vector red; the enzyme was inactivated between the steps by treatment with 0.2M glycine, pH 2.5. Genotypes are described in Fig. 1 legend.

Figure 3: Stainings of third instar imaginal discs. *a*, Dpp expression (green) anterior to the compartment boundary is shown, as marked by the expression of En (red) throughout the posterior compartment (as *hh*). *b*, Wing-blade area stained for Dpp (green) and Myc epitope (red). Arrow indicates a *smo*- clone in the *dpp* expression region. *c*-*e*, Details of disc in *b*. Similar clones within the *dpp* domain were found in six cases among 350 dissected progeny, of which 1/8 were of the correct genotype. Loss of Myc staining represents cells that have lost *smo*- activity, whereas nearby cells expressing high levels of Myc represent the sister clone carrying two copies of the myc construct. In *c*, arrow points to a *smo*- clone within the posterior compartment. The absence of green staining (Dpp) corresponds cell for cell (*e*) with the absence of red staining *smo*-. *f*, *dpp* expression; *g*. as in *f*, except that clones lacking *smo*- and the catalytic subunit of PKA were induced; these express *dpp* ectopically in the compartment owing to the overexpression of *dpp* (Capdevilla J. & Guerrero I. (1994) *EMBO J.* 13:4459-4468).

METHODS: Imaginal discs were dissected in PBS on ice and fixed in PBS 4% paraformaldehyde. For *a*, discs were collected from dpp/LacZ animals and for *c-e*, from a cross of *smo*D16 ck FRT/CyO females with Hs-FLP; FRT dppLacZ/CyO males, the progeny of which were heat-shocked at several stages during larval development. Discs were stained with anti-En, anti-Dpp (Panganiban, G. et al. (1990) *Mol. Cell. Biol.* 10:2669-2677), anti- -galactosidase (Promega) and anti-Myc epitope antibodies or stained for -galactosidase activity using X-gal, following standard procedures.

Figure 4 (panels A-D) show various steps in the cloning of *smoothened.* A: Southern blot of genomic DNA (digested with *Bam*Hl) from different *smo* alleles, hybridized with a subclone from a P1 clone (see *c*). Six P1 clones, covering the area of 21B7-8, the region to which *smo* maps (based on the exclusion from *Df2L(al)* and *Df2L (PMF)*, were screened for the detection of aberrations in *smo* alleles. B: Developmental northern blot hybridized with subclone as in C. C: Restriction map of the *smo* locus. Complementary DNAs for all three transcripts were isolate. The genomic DNA used to make a transgenic fly is indicated. Bold bars indicate the *smo* gene exons. D: The one-letter amino-acid sequence of the open reading frame in the *smo* gene is shown. Three more methionine codons are found in-frame before the indicated start of translation, two of which are not surrounded by residues creating a good translation start site ((Brown, N.H. & Kafatos, F.C. (1988) *J. Mol. Biol.* 203:425-437); a third methionine, just one codon before, is interrupted by an intron. Hydrophobic stretches are underlined, the first one probably representing a signal sequence. Arrow denotes a putative signal peptidase cleavage site. The homology with the *Drosophila Fz* protein is confined to the putative transmembrane region.

METHODS: P1 phage DNA was prepared by the alkaline lysis method and digested; DNA fragments were isolated from agarose gels by spinning through glasswool and subcloned in BlueScript)Stratagene). Genomic fly DNA was prepared, digested with restriction enzymes, separated on agarose gels and transferred to Hybond N (Amersham). P1 fragments were labeled with [32P]dCTP (Amersham protocol) and hybridized to the filters under standard conditions. Total RNA from different stages was prepared by guanidine-HCl extraction and acetic acid precipitation. RNA samples were run under identical conditions and the amount of RNA loaded was controlled by staining the blot with methylene blue after hybridization. The restriction fragment indicated in *c* was isolated and cloned into pCaSpeR for transformation. Transgenic flies were generated and identified among the F1 progeny on the basis of their eye pigmentation, and balanced lines were established. A partial complementary DMA for *smo* was isolated from a 0-2 h cDNA library (Brown, N.H. and Kafatos, F.C. (1988) *supra*) by hybridization with a probe contained within the rescue

fragment. The *smo* cDNA and genomic subclones completely covering the region of the rescue fragment were sequenced using the dideoxy method with Sequenase (USB); both strands of the DNA were sequenced at least twice. The start of transcription and intron-exon boundaries were confirmed by sequencing fragments generated by RT-PCR, using primers designed from the genomic sequence.

Detailed Description of the Invention

Of particular importance in the development and maintenance of tissue in vertebrate animals is a type of extracellular communication called induction, which occurs between neighboring cell layers and tissues. In inductive interactions, chemical signals secreted by one cell population influence the developmental fate of a second cell population. Typically, cells responding to the inductive signals are diverted from one cell fate to another, neither of which is the same as the fate of the signaling cells.

Inductive signals are key regulatory proteins that function in vertebrate pattern formation, and are present in important signaling centers known to operate embryonically, for example, to define the organization of the vertebrate embryo. For example, these signaling structures include the notochord, a transient structure which initiates the formation of the nervous system and helps to define the different types of neurons within it. The notochord also regulates mesodermal patterning along the body axis. Another distinct group of cells having apparent signaling activity is the floorplate of the neural tube (the precursor of the spinal cord and brain) which also signals the differentiation of different nerve cell types. It is also generally believed that the region of mesoderm at the bottom of the buds which form the limbs (called the Zone of Polarizing Activity or ZPA) operates as a signaling center by secreting a morphogen which ultimately produces the correct patterning of the developing limbs.

The regulation of *hedgehog* protein signaling is an important mechanism for developmental control. Members of the *hedgehog* family of secreted proteins control a number of important inductive interactions in the development of both vertebrates and *Drosophila* (Ingham, P.W. (1995) *Opin. Gen. Dev.* 5:492-498). In *Drosophila*, *hedgehog* is one of the segment-polarity genes, mutations of which disrupt the pattern and polarity of individual embryonic segments (Nüsslein-Volhard, C. & Wieschaus, E. (1980) *Nature* 287:795-801) and their adult derivatives (Williams, J.A. & Carroll, S.B. (1993) *Bioessays* 15:567-577).

The present invention concerns the discovery of a new family of cell surface proteins, referred to herein as "*smoothened*" proteins. Here we show that the *smoothened* gene product is required for the response of cells to *hedgehog* signaling during the

development. Sequence analysis of the *smoothened* transcription unit reveals a single open reading frame encoding a protein with seven putative transmembrane domains. This structure is typical of G-protein-coupled receptors. As described herein, the vertebrate *smoothened* proteins exhibit spatially and temporally restricted expression domains indicative of important roles in *hedgehog*-mediated induction.

The sequence of exemplary *smoothened* genes cloned from various metazoan organisms (*c.f.*, Table 1 below) indicates it encodes a receptor-like serpentine protein that may be anchored at the cell membrane. Comparison of *smoothened* sequences from drosophila, chicken, rat and human clones suggests that *smoothened* is an integral membrane proteins with seven membrane spanning α helices, a long cytoplasmic tail, and a conserved signal peptide sequence. Moreover, analysis of the protein sequences suggests potential sites for modification by N-linked glycosylation, as well as potential phosphorylation sites for G-protein coupled receptors and cAMP-dependent kinases (e.g., PKA). The vertebrate *smoothened* proteins also include potential protein-protein interaction modules such as leucine zippers and RGD sequences. The *smoothened* coding sequences has weak homology with the frizzled genes, particularly across the transmembrane domains. However, within the serpentine receptor family, the evidence provided herein suggests that these genes comprise a novel sub-family of receptors.

The *smoothened* proteins, through their ability to associate with *patched* and/or *hedgehog* proteins, are apparently capable of modulating *hedgehog* signaling. The *smoothened* proteins may function as a constitutively active signaling protein whose signal transduction capabilities are inhibited by interaction with *patched*, a *hedgehog* receptor (or subunit thereof). Binding of *hedgehog* to *patched* alters the interaction of *smoothened* and *patched*, and relieves the *patched*-mediated inhibition of *smoothened* signal transduction. Thus, the *smoothened* polypeptides of the present invention may affect a number of *hedgehog*-mediated biological activities including: an ability to modulate proliferation, survival and/or differentiation of mesodermally-derived tissue, such as tissue derived from dorsal mesoderm, cartilage and tissue involved in spermatogenesis; the ability to modulate proliferation, survival and/or differentiation of ectodermally-derived tissue, such as tissue derived from the epidermis, neural tube, neural crest, or head mesenchyme; the ability to modulate proliferation, survival and/or differentiation of endodermally-derived tissue, such as tissue derived from the primitive gut.

As described in the appended examples, a drosophila *smoothened* orf (open reading frame) was identified in a screen for the ability of expressed transgene genomic fragments to rescue a *smo* phenotype. A chicken *smoothened* cDNA was isolated by hybridization under low to medium stringency conditions with the drosophila *smoothened* coding sequence. In addition to the chicken *smoothened* clone, the art has recently reported cDNA

clones from other vertebrates, including human and rodent *smoothened* genes. According to the appended sequence listing, (see also Table 1) a drosophila *smoothened* polypeptide is encoded by SEQ ID No:1; a human *smoothened* polypeptide is encoded by SEQ ID No:2; a rat *smoothened* polypeptide is encoded by SEQ ID No:3; and a chicken *smoothened* polypeptide is encoded by SEQ ID No:4.

Table 1
Guide to *smoothened* sequences in Sequence Listing

	Nucleotide	Amino Acid
Drosophila *smoothened*	SEQ ID No. 1	SEQ ID No. 5
Human *smoothened*	SEQ ID No. 2	SEQ ID No. 6
Rat *smoothened*	SEQ ID No. 3	SEQ ID No. 7
Chicken *smoothened*	SEQ ID No. 4	SEQ ID No. 8

The overall sequence identity between the *smoothened* proteins is shown in Table 2.

Table 2
Amino acid sequence identity between *smoothened* proteins.

	Chicken	Human	Rat	Drosophila
Chicken	-			
Human	64%	-		
Rat	63%	93%	-	
Drosophila	25%	31%	30%	-

It is contemplated by the present invention that the cloned *smoothened* genes set out in the appended sequence listing, in addition to representing an inter-species family of related genes, are also each part of an intra-species family. That is, it is anticipated that other paralogs of the human and mouse *smoothened* proteins exist in those animals in much the same manner as multiple *frizzled* proteins have been identified, and orthologs of each *smoothened* gene are conserved amongst other animals.

In addition to the sequence variation between the various *smoothened* homologs, the vertebrate *smoothened* proteins are apparently present naturally in a number of different forms, including a pro-form. The pro-form includes an N-terminal signal peptide for directed secretion of at least the first extracellular domain of the protein, while the full-length mature form may lack this signal sequence. Further processing of the mature form may also occur in some instances to yield biologically active extracellular or intracellular

fragments of the protein. The *smoothened* proteins may also be modified post-translationally, such as by O-, S- and/or N-linked glycosylation. Potential Asn-glycosylation sites are shown in figure 4.

Smo mutants display phenotypes similar to the *hh* mutants, both in embryos and in clones of mutant cells in *drosophila* imaginal disks. Moreover, the effects of ectopic *hedgehog* become negligible when cells lack *smoothened*. When embryos lack both *smoothened* and *patched*, they have a similar phenotype to *smo* single mutants, indicating that *patched* is located genetically upstream of *smoothened*. It is postulated that secreted *hedgehog* protein binds to *patched*. This binding relieves the *patched*-dependent inhibition of *smoothened*, and may involve direct contact between *smoothened* and *patched*. Once it is relived from inhibitory signals, *smoothened* activates the downstream genes wg (Wnts in vertebrates), dpp (TGFβ proteins) and *patched*, through the signaling components fused, costal-2 and cubitus interruptus (Gli's). This pathway may also involve the inhibition of protein kinase A (PKA), though the latter may merely act in parallel. In this case, *smoothened* would have a constitutive (*hedgehog*-independent) activity in the absence of any inhibition by *patched*.

Accordingly, certain aspects of the present invention relate to nucleic acids encoding *smoothened* polypeptides, the *smoothened* polypeptides themselves (including various fragments), antibodies immunoreactive with *smoothened* proteins, and preparations of such compositions. Moreover, the present invention provides diagnostic and therapeutic assays and reagents for detecting and treating disorders involving, for example, aberrant expression (or loss thereof) of *smoothened*, *smoothened* ligands, or signal transducers thereof.

In addition, drug discovery assays are provided for identifying agents which can modulate the biological function of *smoothened* proteins, such as by altering the interaction of *smoothened* and *patched* proteins, or other extracellular/matrix factors, or the ability of *smoothened* proteins to transduce intracellular signals. Such agents can be useful therapeutically to alter the growth, maintenance and/or differentiation of a tissue, particularly a mesodermally-derived tissue, such cartilage, tissue involved in spermatogenesis and tissue derived from dorsal mesoderm; ectodermally-derived tissue, such as tissue derived from the epidermis, neural tube, neural crest, or head mesenchyme; endodermally-derived tissue, such as tissue derived from the primitive gut. Other aspects of the invention are described below or will be apparent to those skilled in the art in light of the present disclosure.

For convenience, certain terms employed in the specification and appended claims are collected here.

The term "*smoothened*" polypeptide refers to a family of polypeptides characterized at least in part by being identical or sharing a degree of sequence homology with all or a

portion of the a *smoothened* polypeptide represented in any of SEQ ID Nos: 5-8. The *smoothened* polypeptides can be cloned or purified from any of a number of eukaryotic organisms, especially vertebrates, and particularly mammals. Moreover, other *smoothened* polypeptides can be generated according to the present invention, which polypeptides do not ordinarily exist in nature, but rather are generated by non-natural mutagenic techniques.

A "transmembrane " region refers to sequence of amino acids that is located in the cellular membrane, e.g., retained in the membrane at the cell surface.

A "glycosylated" *smoothened* polypeptide is an *smoothened* polypeptide having a covalent linkage with a glycosyl group (e.g. a derivatized with a carbohydrate). For instance, the *smoothened* protein can be glycosylated on an existing residue, or can be mutated to preclude carbohydrate attachment, or can be mutated to provide new glycosylation sites, such as for N-linked or O-linked glycosylation.

As used herein, the term "vertebrate *hedgehog* protein" refers to vertebrate intercellular signaling molecules related to the *Drosophilia hedgehog* protein. Three of the vertebrate *hedgehog* proteins, *Desert hedgehog* (*Dhh*), *Sonic hedgehog* (*Shh*) and *Indian hedgehog* (*Ihh*), apparently exist in all vertebrates, including amphibians, fish, birds, and mammals. Other members of this family, such as *Banded hedgehog*, *Cephalic hedgehog*, *tiggy-winkle hedgehog*, and *echidna hedgehog* have been so far identified in fish and/or amphibians. Exemplary *hedgehog* polypeptides are described in PCT applications WO96/17924, WO96/16668, WO95/18856.

As used herein, the term "nucleic acid" refers to polynucleotides such as deoxyribonucleic acid (DNA), and, where appropriate, ribonucleic acid (RNA). The term should also be understood to include, as equivalents, analogs of either RNA or DNA made from nucleotide analogs, and, as applicable to the embodiment being described, single (sense or antisense) and double-stranded polynucleotides.

As used herein, the term "gene" or "recombinant gene" refers to a nucleic acid comprising an open reading frame encoding a *smoothened* polypeptide, including both exon and (optionally) intron sequences. A "recombinant gene" refers to nucleic acid encoding a *smoothened* polypeptide and comprising *smoothened*-encoding exon sequences, though it may optionally include intron sequences which are derived from, for example, a chromosomal *smoothened* gene (see SEQ ID NO. 9) or from an unrelated chromosomal gene. Exemplary recombinant genes encoding the subject *smoothened* polypeptide are represented in the appended Sequence Listing. The term "intron" refers to a DNA sequence present in a given *smoothened* gene which is not translated into protein and is generally found between exons.

As used herein, the term "transfection" means the introduction of a nucleic acid, e.g., an expression vector, into a recipient cell by nucleic acid-mediated gene transfer. "Transformation", as used herein, refers to a process in which a cell's genotype is changed as a result of the cellular uptake of exogenous DNA or RNA, and, for example, the transformed cell expresses a recombinant form of a *smoothened* polypeptide or, where anti-sense expression occurs from the transferred gene, the expression of a naturally-occurring form of the *smoothened* protein is disrupted.

As used herein, the term "specifically hybridizes" refers to the ability of a nucleic acid probe/primer of the invention to hybridize to at least 15 consecutive nucleotides of a *smoothened* gene, such as a *smoothened* sequence designated in any one or more of SEQ ID Nos: 1-4, or a sequence complementary thereto, or naturally occurring mutants thereof, such that it has less than 15%, preferably less than 10%, and more preferably less than 5% background hybridization to a cellular nucleic acid (e.g., mRNA or genomic DNA) encoding a protein other than a *smoothened* protein, as defined herein.

An "effective amount" of a *hedgehog* polypeptide, or a bioactive fragment thereof, with respect to the subject method of treatment, refers to an amount of agonist or antagonist in a preparation which, when applied as part of a desired dosage regimen, provides modulation of growth, differentiation or survival of cells, e.g., modulation of spermatogenesis, neuronal differentiation, or skeletogenesis, e.g., osteogenesis, chondrogenesis, or limb patterning.

As used herein, "phenotype" refers to the entire physical, biochemical, and physiological makeup of a cell, e.g., having any one trait or any group of traits.

The terms "induction" or "induce", as relating to the biological activity of a *hedgehog* protein, refers generally to the process or act of causing to occur a specific effect on the phenotype of cell. Such effect can be in the form of causing a change in the phenotype, e.g., differentiation to another cell phenotype, or can be in the form of maintaining the cell in a particular cell, e.g., preventing dedifferentation or promoting survival of a cell.

A "patient" or "subject" to be treated can mean either a human or non-human animal.

As used herein, the term "vector" refers to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. One type of preferred vector is an episome, i.e., a nucleic acid capable of extra-chromosomal replication. Preferred vectors are those capable of autonomous replication and/expression of nucleic acids to which they are linked. Vectors capable of directing the expression of genes to which they are operatively linked are referred to herein as "expression vectors". In general, expression

vectors of utility in recombinant DNA techniques are often in the form of "plasmids" which refer generally to circular double stranded DNA loops which, in their vector form are not bound to the chromosome. In the present specification, "plasmid" and "vector" are used interchangeably as the plasmid is the most commonly used form of vector. However, the invention is intended to include such other forms of expression vectors which serve equivalent functions and which become known in the art subsequently hereto.

"Transcriptional regulatory sequence" is a generic term used throughout the specification to refer to DNA sequences, such as initiation signals, enhancers, and promoters, which induce or control transcription of protein coding sequences with which they are operably linked. In preferred embodiments, transcription of a recombinant *smoothened* gene is under the control of a promoter sequence (or other transcriptional regulatory sequence) which controls the expression of the recombinant gene in a cell-type in which expression is intended. It will also be understood that the recombinant gene can be under the control of transcriptional regulatory sequences which are the same or which are different from those sequences which control transcription of the naturally-occurring forms of *smoothened* genes.

As used herein, the term "tissue-specific promoter" means a DNA sequence that serves as a promoter, i.e., regulates expression of a selected DNA sequence operably linked to the promoter, and which effects expression of the selected DNA sequence in specific cells of a tissue, such as cells of neuronal or hematopoietic origin. The term also covers so-called "leaky" promoters, which regulate expression of a selected DNA primarily in one tissue, but can cause at least low level expression in other tissues as well.

As used herein, the term "target tissue" refers to connective tissue, cartilage, bone tissue or limb tissue, which is either present in an animal, e.g., a mammal, e.g., a human or is present in in vitro culture, e.g., a cell culture.

As used herein, a "transgenic animal" is any animal, preferably a non-human mammal, bird or an amphibian, in which one or more of the cells of the animal contain heterologous nucleic acid introduced by way of human intervention, such as by transgenic techniques well known in the art. The nucleic acid is introduced into the cell, directly or indirectly by introduction into a precursor of the cell, by way of deliberate genetic manipulation, such as by microinjection or by infection with a recombinant virus. The term genetic manipulation does not include classical cross-breeding, or *in vitro* fertilization, but rather is directed to the introduction of a recombinant DNA molecule. This molecule may be integrated within a chromosome, or it may be extrachromosomally replicating DNA. In an exemplary transgenic animal, the transgenc causes cells to express a recombinant form of a *smoothened* protein, e.g. either agonistic or antagonistic forms. However, transgenic animals in which the recombinant *smoothened* gene is silent are also contemplated, as for

example, the FLP or CRE recombinase dependent constructs described below. Moreover, "transgenic animal" also includes those recombinant animals in which gene disruption of one or more *smoothened* genes is caused by human intervention, including both recombination and antisense techniques.

The "non-human animals" of the invention include vertebrates such as rodents, non-human primates, livestock, avian species, amphibians, reptiles, etc. The term "chimeric animal" is used herein to refer to animals in which the recombinant gene is found, or in which the recombinant is expressed in some but not all cells of the animal. The term "tissue-specific chimeric animal" indicates that a recombinant *smoothened* gene is present and/or expressed or disrupted in some tissues but not others.

As used herein, the term "transgene" means a nucleic acid sequence (encoding, e.g., a *smoothened* polypeptide, or pending an antisense transcript thereto), which is partly or entirely heterologous, i.e., foreign, to the transgenic animal or cell into which it is introduced, or, is homologous to an endogenous gene of the transgenic animal or cell into which it is introduced, but which is designed to be inserted, or is inserted, into the animal's genome in such a way as to alter the genome of the cell into which it is inserted (e.g., it is inserted at a location which differs from that of the natural gene or its insertion results in a knockout). A transgene can include one or more transcriptional regulatory sequences and any other nucleic acid, such as introns, that may be necessary for optimal expression of a selected nucleic acid.

As is well known, genes for a particular polypeptide may exist in single or multiple copies within the genome of an individual. Such duplicate genes may be identical or may have certain modifications, including nucleotide substitutions, additions or deletions, which all still code for polypeptides having substantially the same activity. The term "DNA sequence encoding a *smoothened* polypeptide" may thus refer to one or more genes within a particular individual. Moreover, certain differences in nucleotide sequences may exist between individuals of the same species, which are called alleles. Such allelic differences may or may not result in differences in amino acid sequence of the encoded polypeptide yet still encode a protein with the same biological activity.

"Homology" refers to sequence similarity between two peptides or between two nucleic acid molecules. Homology can be determined by comparing a position in each sequence which may be aligned for purposes of comparison. When a position in the compared sequence is occupied by the same base or amino acid, then the molecules are homologous at that position. A degree of homology between sequences is a function of the number of matching or homologous positions shared by the sequences. An "unrelated" or "non-homologous" sequence shares less than 40 percent identity, though preferably less than 25 percent identity, with a *smoothened* sequence of the present invention.

The term "ortholog" refers to genes or proteins which are homologs via speciation, e.g., closely related and assumed to have common descent based on structural and functional considerations. Orthologous proteins function as recognizably the same activity in different species. The term "paralog" refers to genes or proteins which are homologs via gene duplication, e.g., duplicated variants of a gene within a genome. See also, Fritch, WM (1970) *Syst Zool* 19:99-113.

"Cells," "host cells" or "recombinant host cells" are terms used interchangeably herein. It is understood that such terms refer not only to the particular subject cell but to the progeny or potential progeny of such a cell. Because certain modifications may occur in succeeding generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term as used herein.

A "chimeric protein" or "fusion protein" is a fusion of a first amino acid sequence encoding a *smoothened* polypeptide with a second amino acid sequence defining a domain (e.g. polypeptide portion) foreign to and not substantially homologous with any domain of a *smoothened* protein. A chimeric protein may present a foreign domain which is found (albeit in a different protein) in an organism which also expresses the first protein, or it may be an "interspecies", "intergenic", etc. fusion of protein structures expressed by different kinds of organisms. In general, a fusion protein can be represented by the general formula X-*smoothened*-Y, wherein *smoothened* represents a portion of the fusion protein which is derived from a *smoothened* protein, and X and Y are, independently, absent or represent amino acid sequences which are not related to a *smoothened* sequences in an organism.

As used herein, a "reporter gene construct" is a nucleic acid that includes a "reporter gene" operatively linked to a transcriptional regulatory sequences. Transcription of the reporter gene is controlled by these sequences. The activity of at least one or more of these control sequences is directly or indirectly regulated by a signal transduction pathway involving a phospholipase, e.g., is directly or indirectly regulated by a second messenger produced by the phospholipase activity. The transcriptional regulatory sequences can include a promoter and other regulatory regions, such as enhancer sequences, that modulate the activity of the promoter, or regulatory sequences that modulate the activity or efficiency of the RNA polymerase that recognizes the promoter, or regulatory sequences that are recognized by effector molecules, including those that are specifically induced upon activation of a phospholipase. For example, modulation of the activity of the promoter may be effected by altering the RNA polymerase binding to the promoter region, or, alternatively, by interfering with initiation of transcription or elongation of the mRNA. Such sequences are herein collectively referred to as transcriptional regulatory elements or sequences. In addition, the construct may include sequences of nucleotides that alter the

stability or rate of translation of the resulting mRNA in response to second messages, thereby altering the amount of reporter gene product.

The term "isolated" as also used herein with respect to nucleic acids, such as DNA or RNA, refers to molecules separated from other DNAs, or RNAs, respectively, that are present in the natural source of the macromolecule. For example, an isolated nucleic acid encoding a *smoothened* polypeptide preferably includes no more than 10 kilobases (kb) of nucleic acid sequence which naturally immediately flanks the *smoothened* gene in genomic DNA, more preferably no more than 5kb of such naturally occurring flanking sequences, and most preferably less than 1.5kb of such naturally occurring flanking sequence. The term isolated as used herein also refers to a nucleic acid or peptide that is substantially free of cellular material, or culture medium when produced by recombinant DNA techniques, or chemical precursors or other chemicals when chemically synthesized. Moreover, an "isolated nucleic acid" is meant to include nucleic acid fragments which are not naturally occurring as fragments and would not be found in the natural state.

As described below, one aspect of the invention pertains to isolated nucleic acids comprising nucleotide sequences encoding *smoothened* polypeptides, and/or equivalents of such nucleic acids. The term nucleic acid as used herein is intended to include fragments as equivalents. The term equivalent is understood to include nucleotide sequences encoding functionally equivalent *smoothened* polypeptides or functionally equivalent peptides having an activity of a *smoothened* protein such as described herein. Equivalent nucleotide sequences will include sequences that differ by one or more nucleotide substitutions, additions or deletions, such as allelic variants; and will, therefore, include sequences that differ from the nucleotide sequence of the *smoothened* coding sequences shown in any one or more of SEQ ID Nos: 1-4 due to the degeneracy of the genetic code. Equivalents will also include nucleotide sequences that hybridize under stringent conditions (i.e., equivalent to about 20-27°C below the melting temperature (T_m) of the DNA duplex formed in about 1M salt) to the nucleotide sequences represented in SEQ ID No: 1-4 or 9. In one embodiment, equivalents will further include nucleic acid sequences derived from and evolutionarily related to, a nucleotide sequences shown in SEQ ID No: 1, SEQ ID No: 2, SEQ ID No: 3 and/or SEQ ID No: 4.

Moreover, it will be generally appreciated that, under certain circumstances, it may be advantageous to provide homologs of a *smoothened* polypeptide which function in a limited capacity as one of either an agonist (e.g., mimics or potentiates a bioactivity of the wild-type *smoothened* protein) or an antagonist (e.g., inhibits a bioactivity of the wild-type *smoothened* protein), in order to promote or inhibit only a subset of the biological activities of the naturally-occurring form of the protein. Thus, specific biological effects can be elicited by treatment with a homolog of limited function. For example, truncated forms of a

smoothened protein, e.g., soluble fragments of the extracellular domain, may competitively inhibit interaction of the wild-type *smoothened* protein with other proteins (such as *patched* or a ligand for *smoothened*).

Homologs of the subject *smoothened* protein can be generated by mutagenesis, such as by discrete point mutation(s), or by truncation. For instance, mutation can give rise to homologs which retain substantially the same, or merely a subset, of the biological activity of the *smoothened* protein from which it was derived. Alternatively, antagonistic forms of the protein can be generated which are able to inhibit the function of the naturally occurring form of the protein, such as by competitively binding to *patched* proteins and competing with wild-type *smoothened*, or binding to other *smoothened* interacting proteins. Thus, the *smoothened* protein and homologs thereof provided by the subject invention may be either positive or negative regulators of cell growth, death and/or differentiation.

In general, polypeptides referred to herein as having an activity of a *smoothened* protein (e.g., are "bioactive") are defined as polypeptides which include an amino acid sequence corresponding (e.g., identical or similar) to all or a portion of the amino acid sequences of the *smoothened* protein shown in SEQ ID No: 5, SEQ ID No: 6, SEQ ID No: 7 or SEQ ID No: 8, and which agonize or antagonize all or a portion of the biological/biochemical activities of a naturally occurring *smoothened* protein. Examples of such biological activity includes the ability to interact with *patched*, (optionally) interact with *hedgehog*, regulate fused, costal-2 or Gli activities. The bioactivity of certain embodiments of the subject *smoothened* polypeptides can be characterized in terms of an ability to promote differentiation and/or maintenance of cells and tissue from mesodermally-derived tissue, such as tissue derived from dorsal mesoderm; ectodermally-origin, such as tissue derived from the neural tube, neural crest, or head mesenchyme; or endodermally-derived tissue, such as tissue derived from the primitive gut.

Other biological activities of the subject *smoothened* proteins are described herein or will be reasonably apparent to those skilled in the art. According to the present invention, a polypeptide has biological activity if it is a specific agonist or antagonist of a naturally-occurring form of a *smoothened* protein.

Preferred nucleic acids encode a *smoothened* polypeptide comprising an amino acid sequence at least 60%, 63%, 70% or 80% homologous, more preferably at least 85% homologous and most preferably at least 93% or 95% homologous with an amino acid sequence of a naturally occurring *smoothened* protein, e.g., such as represented in SEQ ID No: 5, SEQ ID No: 6, SEQ ID No: 7 or SEQ ID No: 8. Nucleic acids which encode polypeptides at least about 98-99% homology with an amino acid sequence represented in SEQ ID No: 5, SEQ ID No: 6, SEQ ID No: 7 or SEQ ID No: 8 are of course also within the scope of the invention, as are nucleic acids identical in sequence with the enumerated

smoothened sequence of the Sequence listing. In one embodiment, the nucleic acid is a cDNA encoding a polypeptide having at least one activity of the subject *smoothened* polypeptide.

In certain preferred embodiments, the invention features a purified or recombinant *smoothened* polypeptide. It will be understood that the *smoothened* protein can include certain post-translational modifications, e.g., glycosylation, phosphorylation and the like, and cleavage of certain sequences, such as pro-sequences.

Another aspect of the invention provides a nucleic acid which hybridizes under high or low stringency conditions to one or more of the nucleic acids represented by SEQ ID Nos: 1-4 and 9. Appropriate stringency conditions which promote DNA hybridization, for example, 6.0 x sodium chloride/sodium citrate (SSC) at about 45°C, followed by a wash of 2.0 x SSC at 50°C, are known to those skilled in the art or can be found in *Current Protocols in Molecular Biology*, John Wiley & Sons, N.Y. (1989), 6.3.1-6.3.6. For example, the salt concentration in the wash step can be selected from a low stringency of about 2.0 x SSC at 50°C to a high stringency of about 0.2 x SSC at 50°C. In addition, the temperature in the wash step can be increased from low stringency conditions at room temperature, about 22°C, to high stringency conditions at about 65°C.

Nucleic acids, having a sequence that differs from the nucleotide sequences shown in any of SEQ ID Nos: 1-4 due to degeneracy in the genetic code are also within the scope of the invention. Such nucleic acids encode functionally equivalent peptides (i.e., a peptide having a biological activity of a *smoothened* polypeptide) but differ in sequence from the sequence shown in the sequence listing due to degeneracy in the genetic code. For example, a number of amino acids are designated by more than one triplet. Codons that specify the same amino acid, or synonyms (for example, CAU and CAC each encode histidine) may result in "silent" mutations which do not affect the amino acid sequence of a *smoothened* polypeptide. However, it is expected that DNA sequence polymorphisms that do lead to changes in the amino acid sequences of the subject *smoothened* polypeptides will exist among, for example, humans. One skilled in the art will appreciate that these variations in one or more nucleotides (up to about 3-5% of the nucleotides) of the nucleic acids encoding polypeptides having an activity of a *smoothened* polypeptide may exist among individuals of a given species due to natural allelic variation.

As used herein, a *smoothened* gene fragment refers to a nucleic acid having fewer nucleotides than the nucleotide sequence encoding the entire mature form of a *smoothened* protein yet which (preferably) encodes a polypeptide which retains some biological activity of the full length protein. Fragment sizes contemplated by the present invention include, for example, 5, 10, 25, 50, 75, 100, or 200 amino acids in length. In a preferred embodiment of a truncated receptor, the polypeptide will include all or a sufficient portion of a *patched-*

interacting domain or, as appropriate, a ligand domain or intracellular domain involved in signal transduction.

As indicated by the examples set out below, *smoothened* protein-encoding nucleic acids can be obtained from mRNA present in cells of metazoan organisms. As further illustrated in the examples, it is also possible to obtain nucleic acids encoding *smoothened* polypeptides of the present invention from genomic DNA from both adults and embryos. For example, a gene encoding a *smoothened* protein can be cloned from either a cDNA or a genomic library in accordance with protocols described herein, as well as those generally known to persons skilled in the art. A cDNA encoding a *smoothened* protein can be obtained by isolating total mRNA from a cell, such as a mammalian cell, e.g. a human cell, as desired. Double stranded cDNAs can be prepared from the total mRNA, and subsequently inserted into a suitable plasmid or bacteriophage vector using any one of a number of known techniques. The gene encoding a *smoothened* protein can also be cloned using established polymerase chain reaction techniques in accordance with the nucleotide sequence information provided by the invention. The nucleic acid of the invention can be DNA or RNA. A preferred nucleic acid is a cDNA including a nucleotide sequence represented by any one of SEQ ID No: 1, SEQ ID No: 2, SEQ ID No: 3, or SEQ ID No: 4.

Another aspect of the invention relates to the use of the isolated nucleic acid in "antisense" therapy. As used herein, "antisense" therapy refers to administration or *in situ* generation of oligonucleotide probes or their derivatives which specifically hybridize (e.g. binds) under cellular conditions, with the cellular mRNA and/or genomic DNA encoding a subject *smoothened* protein so as to inhibit expression of that protein, e.g. by inhibiting transcription and/or translation. The binding may be by conventional base pair complementarity, or, for example, in the case of binding to DNA duplexes, through specific interactions in the major groove of the double helix. In general, "antisense" therapy refers to the range of techniques generally employed in the art, and includes any therapy which relies on specific binding to oligonucleotide sequences.

An antisense construct of the present invention can be delivered, for example, as an expression plasmid which, when transcribed in the cell, produces RNA which is complementary to at least a unique portion of the cellular mRNA which encodes a *smoothened* protein. Alternatively, the antisense construct is an oligonucleotide probe which is generated *ex vivo* and which, when introduced into the cell causes inhibition of expression by hybridizing with the mRNA and/or genomic sequences of a *smoothened* gene. Such oligonucleotide probes are preferably modified oligonucleotides which are resistant to endogenous nucleases, e.g. exonucleases and/or endonucleases, and are therefore stable *in vivo*. Exemplary nucleic acid molecules for use as antisense oligonucleotides are phosphoramidate, phosphothioate and methylphosphonate analogs of DNA (see also U.S.

Patents 5,176,996; 5,264,564; and 5,256,775), or peptide nucleic acids (PNAs). Additionally, general approaches to constructing oligomers useful in antisense therapy have been reviewed, for example, by Van der Krol et al. (1988) Biotechniques 6:958-976; and Stein et al. (1988) Cancer Res 48:2659-2668.

Accordingly, the modified oligomers of the invention are useful in therapeutic, diagnostic, and research contexts. In therapeutic applications, the oligomers are utilized in a manner appropriate for antisense therapy in general. For such therapy, the oligomers of the invention can be formulated for a variety of routes of administration, including systemic and topical or localized administration. Techniques and formulations generally may be found in Remmington's Pharmaceutical Sciences, Meade Publishing Co., Easton, PA. For systemic administration, injection is preferred, including intramuscular, intravenous, intraperitoneal, and subcutaneous. For injection, the oligomers of the invention can be formulated in liquid solutions, preferably in physiologically compatible buffers such as Hank's solution or Ringer's solution. In addition, the oligomers may be formulated in solid form and redissolved or suspended immediately prior to use. Lyophilized forms are also included.

Systemic administration can also be by transmucosal or transdermal means, or the compounds can be administered orally. For transmucosal or transdermal administration, penetrants appropriate to the barrier to be permeated are used in the formulation. Such penetrants are generally known in the art, and include, for example, for transmucosal administration bile salts and fusidic acid derivatives. In addition, detergents may be used to facilitate permeation. Transmucosal administration may be through nasal sprays or using suppositories. For oral administration, the oligomers are formulated into conventional oral administration forms such as capsules, tablets, and tonics. For topical administration, the oligomers of the invention are formulated into ointments, salves, gels, or creams as generally known in the art.

In addition to use in therapy, the oligomers of the invention may be used as diagnostic reagents to detect the presence or absence of the target DNA or RNA sequences to which they specifically bind. Such diagnostic tests are described in further detail below.

Likewise, the antisense constructs of the present invention, by antagonizing the normal biological activity of a *smoothened* protein, e.g., by reducing the level of its expression, can be used in the manipulation of tissue, e.g. tissue maintenance, differentiation or growth, both *in vivo* and *ex vivo*.

Furthermore, the anti-sense techniques (e.g. microinjection of antisense molecules, or transfection with plasmids whose transcripts are anti-sense with regard to a *smoothened* mRNA or gene sequence) can be used to investigate the role of *smoothened* in developmental events, as well as the normal cellular function of *smoothened* in adult tissue.

Such techniques can be utilized in cell culture, but can also be used in the creation of transgenic animals (described *infra*).

This invention also provides expression vectors containing a nucleic acid encoding a *smoothened* polypeptide, operably linked to at least one transcriptional regulatory sequence. Operably linked is intended to mean that the nucleotide sequence is linked to a regulatory sequence in a manner which allows expression of the nucleotide sequence. Regulatory sequences are art-recognized and are selected to direct expression of the subject *smoothened* proteins. Accordingly, the term transcriptional regulatory sequence includes promoters, enhancers and other expression control elements. Such regulatory sequences are described in Goeddel; Gene Expression Technology: Methods in Enzymology 185, Academic Press, San Diego, CA (1990). For instance, any of a wide variety of expression control sequences, sequences that control the expression of a DNA sequence when operatively linked to it, may be used in these vectors to express DNA sequences encoding *smoothened* polypeptides of this invention. Such useful expression control sequences, include, for example, a viral LTR, such as the LTR of the Moloney murine leukemia virus, the early and late promoters of SV40, adenovirus or cytomegalovirus immediate early promoter, the lac system, the trp system, the TAC or TRC system, T7 promoter whose expression is directed by T7 RNA polymerase, the major operator and promoter regions of phage λ, the control regions for fd coat protein, the promoter for 3-phosphoglycerate kinase or other glycolytic enzymes, the promoters of acid phosphatase, e.g., Pho5, the promoters of the yeast α-mating factors, the polyhedron promoter of the baculovirus system and other sequences known to control the expression of genes of prokaryotic or eukaryotic cells or their viruses, and various combinations thereof. It should be understood that the design of the expression vector may depend on such factors as the choice of the host cell to be transformed and/or the type of protein desired to be expressed.

Moreover, the vector's copy number, the ability to control that copy number and the expression of any other proteins encoded by the vector, such as antibiotic markers, should also be considered. In one embodiment, the expression vector includes a recombinant gene encoding a polypeptide having an agonistic activity of a subject *smoothened* polypeptide, or alternatively, encoding a polypeptide which is an antagonistic form of the *smoothened* protein. Such expression vectors can be used to transfect cells and thereby produce polypeptides, including fusion proteins, encoded by nucleic acids as described herein.

Moreover, the gene constructs of the present invention can also be used as a part of a gene therapy protocol to deliver nucleic acids, e.g., encoding either an agonistic or antagonistic form of a subject *smoothened* proteins or an antisense molecule described above. Thus, another aspect of the invention features expression vectors for *in vivo* or *in vitro* transfection and expression of a *smoothened* polypeptide or antisense molecule in

particular cell types so as to reconstitute the function of, or alternatively, abrogate all or a portion of the biological function of *smoothened*-induced transcription in a tissue in which the naturally-occurring form of the protein is misexpressed (or has been disrupted); or to deliver a form of the protein which alters proliferation, maintenance or differentiation of tissue, or which inhibits neoplastic or hyperplastic proliferation.

Expression constructs of the subject *smoothened* polypeptides, as well as antisense constructs, may be administered in any biologically effective carrier, e.g. any formulation or composition capable of effectively delivering the recombinant gene to cells *in vivo*. Approaches include insertion of the subject gene in viral vectors including recombinant retroviruses, adenovirus, adeno-associated virus, and herpes simplex virus-1, or recombinant bacterial or eukaryotic plasmids. Viral vectors transfect cells directly; plasmid DNA can be delivered with the help of, for example, cationic liposomes (lipofectin) or derivatized (e.g. antibody conjugated), polylysine conjugates, gramacidin S, artificial viral envelopes or other such intracellular carriers, as well as direct injection of the gene construct or $CaPO_4$ precipitation carried out *in vivo*. It will be appreciated that because transduction of appropriate target cells represents the critical first step in gene therapy, choice of the particular gene delivery system will depend on such factors as the phenotype of the intended target and the route of administration, e.g. locally or systemically. Furthermore, it will be recognized that the particular gene construct provided for *in vivo* transduction of *smoothened* expression are also useful for *in vitro* transduction of cells, such as for use in the *ex vivo* tissue culture systems described below.

A preferred approach for *in vivo* introduction of nucleic acid into a cell is by use of a viral vector containing nucleic acid, e.g. a cDNA encoding the particular *smoothened* polypeptide desired. Infection of cells with a viral vector has the advantage that a large proportion of the targeted cells can receive the nucleic acid. Additionally, molecules encoded within the viral vector, e.g., by a cDNA contained in the viral vector, are expressed efficiently in cells which have taken up viral vector nucleic acid. Retrovirus vectors, adenovirus vectors, adeno-associated and herpes-based virus vectors are exemplary recombinant gene delivery system for the transfer of exogenous genes *in vivo*, particularly into humans. These vectors provide efficient delivery of genes into cells, and the transferred nucleic acids are stably integrated into the chromosomal DNA of the host.

In addition to viral transfer methods, such as those illustrated above, non-viral methods can also be employed to cause expression of a subject *smoothened* polypeptide in the tissue of an animal. Most nonviral methods of gene transfer rely on normal mechanisms used by mammalian cells for the uptake and intracellular transport of macromolecules. In preferred embodiments, non-viral gene delivery systems of the present invention rely on endocytic pathways for the uptake of the subject *smoothened* genes by the targeted cell.

Exemplary gene delivery systems of this type include liposomal derived systems, poly-lysine conjugates, and artificial viral envelopes.

In clinical settings, the gene delivery systems for the therapeutic *smoothened* gene can be introduced into a patient-animal by any of a number of methods, each of which is familiar in the art. For instance, a pharmaceutical preparation of the gene delivery system can be introduced systemically, e.g. by intravenous injection, and specific transduction of the protein in the target cells occurs predominantly from specificity of transfection provided by the gene delivery vehicle, cell-type or tissue-type expression due to the transcriptional regulatory sequences controlling expression of the receptor gene, or a combination thereof. In other embodiments, initial delivery of the recombinant gene is more limited with introduction into the animal being quite localized. For example, the gene delivery vehicle can be introduced by catheter (see U.S. Patent 5,328,470) or by stereotactic injection (e.g. Chen et al. (1994) PNAS 91: 3054-3057). A *smoothened* gene can be delivered in a gene therapy construct by electroporation using techniques described, for example, by Dev et al. ((1994) Cancer Treat Rev 20:105-115).

The pharmaceutical preparation of the gene therapy construct can consist essentially of the gene delivery system in an acceptable diluent, or can comprise a slow release matrix in which the gene delivery vehicle is imbedded. Alternatively, where the complete gene delivery system can be produced intact from recombinant cells, e.g. retroviral vectors, the pharmaceutical preparation can comprise one or more cells which produce the gene delivery system.

In yet another embodiment, the subject invention provides a "gene activation" construct which, by homologous recombination with a genomic DNA, alters the transcriptional regulatory sequences of an endogenous *smoothened* gene. For instance, the gene activation construct can replace the endogenous promoter of a *smoothened* gene with a heterologous promoter, e.g., one which causes constitutive expression of the *smoothened* gene or which causes inducible expression of the gene under conditions different from the normal expression pattern of *smoothened*. A variety of different formats for the gene activation constructs are available. See, for example, the Transkaryotic Therapies, Inc PCT publications WO93/09222, WO95/31560, WO96/29411, WO95/31560 and WO94/12650.

In preferred embodiments, the nucleotide sequence used as the gene activation construct can be comprised of (1) DNA from some portion of the endogenous *smoothened* gene (exon sequence, intron sequence, promoter sequences, etc.) which direct recombination and (2) heterologous transcriptional regulatory sequence(s) which is to be operably linked to the coding sequence for the genomic *smoothened* gene upon recombination of the gene activation construct. For use in generating cultures of

smoothened producing cells, the construct may further include a reporter gene to detect the presence of the knockout construct in the cell.

The gene activation construct is inserted into a cell, and integrates with the genomic DNA of the cell in such a position so as to provide the heterologous regulatory sequences in operative association with the native *smoothened* gene. Such insertion occurs by homologous recombination, i.e., recombination regions of the activation construct that are homologous to the endogenous *smoothened* gene sequence hybridize to the genomic DNA and recombine with the genomic sequences so that the construct is incorporated into the corresponding position of the genomic DNA.

The terms "recombination region" or "targeting sequence" refer to a segment (i.e., a portion) of a gene activation construct having a sequence that is substantially identical to or substantially complementary to a genomic gene sequence, e.g., including 5' flanking sequences of the genomic gene, and can facilitate homologous recombination between the genomic sequence and the targeting transgene construct.

As used herein, the term "replacement region" refers to a portion of a activation construct which becomes integrated into an endogenous chromosomal location following homologous recombination between a recombination region and a genomic sequence.

The heterologous regulatory sequences, e.g., which are provided in the replacement region, can include one or more of a variety elements, including: promoters (such as constitutive or inducible promoters), enhancers, negative regulatory elements, locus control regions, transcription factor binding sites, or combinations thereof. Promoters/enhancers which may be used to control the expression of the targeted gene *in vivo* include, but are not limited to, the cytomegalovirus (CMV) promoter/enhancer (Karasuyama et al., 1989, *J. Exp. Med.*, 169:13), the human β-actin promoter (Gunning et al. (1987) *PNAS* 84:4831-4835), the glucocorticoid-inducible promoter present in the mouse mammary tumor virus long terminal repeat (MMTV LTR) (Klessig et al. (1984) *Mol. Cell Biol.* 4:1354-1362), the long terminal repeat sequences of Moloney murine leukemia virus (MuLV LTR) (Weiss et al. (1985) *RNA Tumor Viruses*, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York), the SV40 early or late region promoter (Bernoist et al. (1981) *Nature* 290:304-310; Templeton et al. (1984) *Mol. Cell Biol.*, 4:817; and Sprague et al. (1983) *J. Virol.*, 45:773), the promoter contained in the 3' long terminal repeat of Rous sarcoma virus (RSV) (Yamamoto et al., 1980, *Cell*, 22:787-797), the herpes simplex virus (HSV) thymidine kinase promoter/enhancer (Wagner et al. (1981) *PNAS* 82:3567-71), and the herpes simplex virus LAT promoter (Wolfe et al. (1992) *Nature Genetics*, 1:379-384).

In still other embodiments, the replacement region merely deletes a negative transcriptional control element of the native gene, e.g., to activate expression, or ablates a positive control element, e.g., to inhibit expression of the targeted gene.

Another aspect of the present invention concerns recombinant forms of the *smoothened* proteins. Recombinant polypeptides preferred by the present invention, in addition to native *smoothened* proteins, are at least 60%, 63%, 64% or 70% homologous, more preferably at least 80% homologous and most preferably at least 85% homologous with an amino acid sequence represented by one or more of SEQ ID Nos: 5, SEQ ID No: 6, SEQ ID No: 7 and SEQ ID No: 8. Polypeptides which possess an activity of a *smoothened* protein (i.e. either agonistic or antagonistic), and which are at least 90%, more preferably at least 93% or 95%, and most preferably at least about 98-99% homologous with SEQ ID No: 5, SEQ ID No: 6, SEQ ID No: 7 and/or SEQ ID No: 8 are also within the scope of the invention. Such polypeptides, as described above, include various truncated forms of the protein.

The term "recombinant *smoothened* polypeptide" refers to a polypeptide which is produced by recombinant DNA techniques, wherein generally, DNA encoding a *smoothened* polypeptide is inserted into a suitable expression vector which is in turn used to transform a host cell to produce the heterologous protein. Moreover, the phrase "derived from", with respect to a recombinant *smoothened* gene, is meant to include within the meaning of "recombinant protein" those proteins having an amino acid sequence of a native *smoothened* protein, or an amino acid sequence similar thereto which is generated by mutations including substitutions and deletions (including truncation) of a naturally occurring form of the protein.

The present invention further pertains to recombinant forms of the subject *smoothened* polypeptides which are encoded by genes derived from a mammal (e.g. a human), reptile or amphibian and which have amino acid sequences evolutionarily related to the *smoothened* protein represented in SEQ ID No: 5, SEQ ID No: 6, SEQ ID No: 7 and SEQ ID No: 8. Such recombinant *smoothened* polypeptides preferably are capable of functioning in one of either role of an agonist or antagonist of at least one biological activity of a wild-type ("authentic") *smoothened* protein of the appended sequence listing.

The present invention also provides methods of producing the subject *smoothened* polypeptides. For example, a host cell transfected with a nucleic acid vector directing expression of a nucleotide sequence encoding the subject polypeptides can be cultured under appropriate conditions to allow expression of the polypeptide to occur. A cell culture includes host cells, media and other byproducts. Suitable media for cell culture are well known in the art. The recombinant *smoothened* polypeptide can be isolated from cell culture medium, host cells, or both using techniques known in the art for purifying proteins including ion-exchange chromatography, gel filtration chromatography, ultrafiltration, electrophoresis, and immunoaffinity purification with antibodies specific for such peptide. In a preferred embodiment, the recombinant *smoothened* polypeptide is isolated in a

membrane fraction of a liposome. For soluble fragments of the protein, the fusion protein can include a domain which facilitates its purification, such as GST fusion protein or poly(His) fusion protein.

This invention also pertains to a host cell transfected to express recombinant forms of the subject *smoothened* polypeptides. The host cell may be any eukaryotic or prokaryotic cell, though eukaryotic cells are preferred, especially mammalian cells. Thus, a nucleotide sequence derived from the cloning of *smoothened* genes, encoding all or a selected portion of a full-length protein, can be used to produce a recombinant form of a *smoothened* polypeptide via microbial or eukaryotic cellular processes. Ligating the polynucleotide sequence into a gene construct, such as an expression vector, and transforming or transfecting into hosts, either eukaryotic (yeast, avian, insect or mammalian) or prokaryotic (bacterial cells), are standard procedures used in producing other well-known proteins, e.g. *patched,* G protein coupled receptors, as well as a wide range of other transmembrane proteins. Similar procedures, or modifications thereof, can be employed to prepare recombinant *smoothened* polypeptides by microbial means or tissue-culture technology in accord with the subject invention.

The recombinant *smoothened* genes can be produced by ligating nucleic acid encoding a *smoothened* polypeptide into a vector suitable for expression in either prokaryotic cells, eukaryotic cells, or both. Expression vectors for production of recombinant forms of the subject *smoothened* polypeptides include plasmids and other vectors. For instance, suitable vectors for the expression of a *smoothened* polypeptide include plasmids of the types: pBR322-derived plasmids, pEMBL-derived plasmids, pEX-derived plasmids, pBTac-derived plasmids and pUC-derived plasmids for expression in prokaryotic cells, such as E. coli.

A number of vectors exist for the expression of recombinant proteins in yeast. For instance, YEP24, YIP5, YEP51, YEP52, pYES2, and YRP17 are cloning and expression vehicles useful in the introduction of genetic constructs into S. cerevisiae (see, for example, Broach et al. (1983) in Experimental Manipulation of Gene Expression, ed. M. Inouye Academic Press, p. 83, incorporated by reference herein). These vectors can replicate in E. coli due the presence of the pBR322 ori, and in S. cerevisiae due to the replication determinant of the yeast 2 micron plasmid. In addition, drug resistance markers such as ampicillin can be used. In an illustrative embodiment, a *smoothened* polypeptide is produced recombinantly utilizing an expression vector generated by sub-cloning the coding sequence of a *smoothened* gene represented in SEQ ID No: 1, SEQ ID No: 2, SEQ ID No: 3, SEQ ID No: 4, or SEQ ID No: 9.

The preferred mammalian expression vectors contain both prokaryotic sequences, to facilitate the propagation of the vector in bacteria, and one or more eukaryotic transcription

units that are expressed in eukaryotic cells. The pcDNAI/amp, pcDNAI/neo, pRc/CMV, pSV2gpt, pSV2neo, pSV2-dhfr, pTk2, pRSVneo, pMSG, pSVT7, pko-neo and pHyg derived vectors are examples of mammalian expression vectors suitable for transfection of eukaryotic cells. Some of these vectors are modified with sequences from bacterial plasmids, such as pBR322, to facilitate replication and drug resistance selection in both prokaryotic and eukaryotic cells. Alternatively, derivatives of viruses such as the bovine papillomavirus (BPV-1), or Epstein-Barr virus (pHEBo, pREP-derived and p205) can be used for transient expression of proteins in eukaryotic cells. The various methods employed in the preparation of the plasmids and transformation of host organisms are well known in the art. For other suitable expression systems for both prokaryotic and eukaryotic cells, as well as general recombinant procedures, see Molecular Cloning A Laboratory Manual, 2nd Ed., ed. by Sambrook, Fritsch and Maniatis (Cold Spring Harbor Laboratory Press: 1989) Chapters 16 and 17.

In some instances, it may be desirable to express the recombinant *smoothened* polypeptide by the use of a baculovirus expression system. Examples of such baculovirus expression systems include pVL-derived vectors (such as pVL1392, pVL1393 and pVL941), pAcUW-derived vectors (such as pAcUW1), and pBlueBac-derived vectors (such as the ß-gal containing pBlueBac III).

When it is desirable to express only a portion of a *smoothened* protein, such as a form lacking a portion of the N-terminus, i.e. a truncation mutant which lacks the signal peptide, it may be necessary to add a start codon (ATG) to the oligonucleotide fragment containing the desired sequence to be expressed. It is well known in the art that a methionine at the N-terminal position can be enzymatically cleaved by the use of the enzyme methionine aminopeptidase (MAP). MAP has been cloned from E. coli (Ben-Bassat et al. (1987) J. Bacteriol. 169:751-757) and Salmonella typhimurium and its *in vitro* activity has been demonstrated on recombinant proteins (Miller et al. (1987) PNAS 84:2718-1722). Therefore, removal of an N-terminal methionine, if desired, can be achieved either *in vivo* by expressing *smoothened*-derived polypeptides in a host which produces MAP (e.g., E. coli or CM89 or S. cerevisiae), or *in vitro* by use of purified MAP (e.g., procedure of Miller et al., supra).

Alternatively, the coding sequences for the polypeptide can be incorporated as a part of a fusion gene including a nucleotide sequence encoding a different polypeptide. This type of expression system can be useful under conditions where it is desirable to produce an immunogenic fragment of a *smoothened* protein. For example, the VP6 capsid protein of rotavirus can be used as an immunologic carrier protein for portions of the *smoothened* polypeptide, either in the monomeric form or in the form of a viral particle. The nucleic acid sequences corresponding to the portion of a subject *smoothened* protein to which

antibodies are to be raised can be incorporated into a fusion gene construct which includes coding sequences for a late vaccinia virus structural protein to produce a set of recombinant viruses expressing fusion proteins comprising *smoothened* epitopes as part of the virion. It has been demonstrated with the use of immunogenic fusion proteins utilizing the Hepatitis B surface antigen fusion proteins that recombinant Hepatitis B virions can be utilized in this role as well. Similarly, chimeric constructs coding for fusion proteins containing a portion of a *smoothened* protein and the poliovirus capsid protein can be created to enhance immunogenicity of the set of polypeptide antigens (see, for example, EP Publication No: 0259149; and Evans et al. (1989) Nature 339:385; Huang et al. (1988) J. Virol. 62:3855; and Schlienger et al. (1992) J. Virol. 66:2).

The Multiple Antigen Peptide system for peptide-based immunization can also be utilized to generate an immunogen, wherein a desired portion of a *smoothened* polypeptide is obtained directly from organo-chemical synthesis of the peptide onto an oligomeric branching lysine core (see, for example, Posnett et al. (1988) JBC 263:1719 and Nardelli et al. (1992) J. Immunol. 148:914). Antigenic determinants of *smoothened* proteins can also be expressed and presented by bacterial cells.

In addition to utilizing fusion proteins to enhance immunogenicity, it is widely appreciated that fusion proteins can also facilitate the expression of proteins, and accordingly, can be used in the expression of the *smoothened* polypeptides of the present invention, particularly truncated forms of the *smoothened* protein. For example, soluble forms of *smoothened* polypeptides can be generated as glutathione-S-transferase (GST-fusion) proteins. Such GST-fusion proteins can enable easy purification of the *smoothened* polypeptide, as for example by the use of glutathione-derivatized matrices (see, for example, Current Protocols in Molecular Biology, eds. Ausubel et al. (N.Y.: John Wiley & Sons, 1991)).

Techniques for making fusion genes are known to those skilled in the art. Essentially, the joining of various DNA fragments coding for different polypeptide sequences is performed in accordance with conventional techniques, employing blunt-ended or stagger-ended termini for ligation, restriction enzyme digestion to provide for appropriate termini, filling-in of cohesive ends as appropriate, alkaline phosphatase treatment to avoid undesirable joining, and enzymatic ligation. In another embodiment, the fusion gene can be synthesized by conventional techniques including automated DNA synthesizers. Alternatively, PCR amplification of gene fragments can be carried out using anchor primers which give rise to complementary overhangs between two consecutive gene fragments which can subsequently be annealed to generate a chimeric gene sequence (see, for example, Current Protocols in Molecular Biology, eds. Ausubel et al. John Wiley & Sons: 1992).

The *smoothened* polypeptides may also be chemically modified to create *smoothened* derivatives by forming covalent or aggregate conjugates with other chemical moieties, such as glycosyl groups, lipids, phosphate, acetyl groups and the like. Covalent derivatives of *smoothened* proteins can be prepared by linking the chemical moieties to functional groups on amino acid sidechains of the protein or at the N-terminus or at the C-terminus of the polypeptide.

The present invention also makes available isolated *smoothened* polypeptides which are isolated from, or otherwise substantially free of other cellular proteins, especially receptors and/or other inductive polypeptides which may normally be associated with the *smoothened* polypeptide. The term "substantially free of other cellular proteins" (also referred to herein as "contaminating proteins") or "substantially pure or purified preparations" are defined as encompassing preparations of *smoothened* polypeptides having less than 20% (by dry weight) contaminating protein, and preferably having less than 5% contaminating protein. Functional forms of the subject polypeptides can be prepared, for the first time, as purified preparations by using a cloned gene as described herein. By "purified", it is meant, when referring to a peptide or DNA or RNA sequence, that the indicated molecule is present in the substantial absence of other biological macromolecules, such as other proteins. The term "purified" as used herein preferably means at least 80% by dry weight, more preferably in the range of 95-99% by weight, and most preferably at least 99.8% by weight, of biological macromolecules of the same type present (but water, buffers, and other small molecules, especially molecules having a molecular weight of less than 5000, can be present). The term "pure" as used herein preferably has the same numerical limits as "purified" immediately above. "Isolated" and "purified" do not encompass either natural materials in their native state or natural materials that have been separated into components (e.g., in an acrylamide gel) but not obtained either as pure (e.g. lacking contaminating proteins, or chromatography reagents such as denaturing agents and polymers, e.g. acrylamide or agarose) substances or solutions. In preferred embodiments, purified *smoothened* preparations will lack any contaminating proteins from the same animal from that *smoothened* is normally produced, as can be accomplished by recombinant expression of, for example, a mammalian *smoothened* protein in a yeast or bacterial cell.

As described above for recombinant polypeptides, isolated *smoothened* polypeptides can include all or a portion of an amino acid sequences corresponding to a *smoothened* polypeptide represented in SEQ ID No: 5, SEQ ID No: 6, SEQ ID No: 7 and SEQ ID No: 8 or homologous sequences thereto.

Isolated peptidyl portions of *smoothened* proteins can also be obtained by screening peptides recombinantly produced from the corresponding fragment of the nucleic acid encoding such peptides. In addition, fragments can be chemically synthesized using

techniques known in the art such as conventional Merrifield solid phase f-Moc or t-Boc chemistry. For example, a *smoothened* polypeptide of the present invention may be arbitrarily divided into fragments of desired length with no overlap of the fragments, or preferably divided into overlapping fragments of a desired length. The fragments can be produced (recombinantly or by chemical synthesis) and tested to identify those peptidyl fragments which can function as either agonists or antagonists of a wild-type (e.g., "authentic") *smoothened* protein, such as by binding to *patched.* For example, Román et al. (1994) *Eur J Biochem* 222:65-73 describe the use of competitive-binding assays using short, overlapping synthetic peptides from larger proteins to identify binding domains.

Modification of the structure of the subject *smoothened* polypeptides can be for such purposes as enhancing therapeutic or prophylactic efficacy, stability (e.g., *ex vivo* shelf life and resistance to proteolytic degradation *in vivo*), or post-translational modifications. Such modified peptides, when designed to retain at least one activity of the naturally-occurring form of the protein, or to produce specific antagonists thereof, are considered functional equivalents of the *smoothened* polypeptides (though they may be agonistic or antagonistic of the bioactivities of the authentic protein). Such modified peptides can be produced, for instance, by amino acid substitution, deletion, or addition.

For example, it is reasonable to expect that an isolated replacement of a leucine with an isoleucine or valine, an aspartate with a glutamate, a threonine with a serine, or a similar replacement of an amino acid with a structurally related amino acid (i.e. isosteric and/or isoelectric mutations) will not have a major effect on the biological activity of the resulting molecule. Conservative replacements are those that take place within a family of amino acids that are related in their side chains. Genetically encoded amino acids are can be divided into four families: (1) acidic = aspartate, glutamate; (2) basic = lysine, arginine, histidine; (3) nonpolar = alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine, tryptophan; and (4) uncharged polar = glycine, asparagine, glutamine, cysteine, serine, threonine, tyrosine. Phenylalanine, tryptophan, and tyrosine are sometimes classified jointly as aromatic amino acids. In similar fashion, the amino acid repertoire can be grouped as (1) acidic = aspartate, glutamate; (2) basic = lysine, arginine histidine, (3) aliphatic = glycine, alanine, valine, leucine, isoleucine, serine, threonine, with serine and threonine optionally be grouped separately as aliphatic-hydroxyl; (4) aromatic = phenylalanine, tyrosine, tryptophan; (5) amide = asparagine, glutamine; and (6) sulfur - containing = cysteine and methionine. (see, for example, Biochemistry, 2nd ed., Ed. by L. Stryer, WH Freeman and Co.: 1981). Whether a change in the amino acid sequence of a peptide results in a functional *smoothened* homolog (e.g. functional in the sense that the resulting polypeptide mimics or antagonizes the authentic form) can be readily determined by assessing the ability of the variant peptide to produce a response in cells in a fashion

similar to the wild-type protein, or competitively inhibit such a response. Polypeptides in which more than one replacement has taken place can readily be tested in the same manner.

This invention further contemplates a method for generating sets of combinatorial point mutants of the subject *smoothened* proteins as well as truncation mutants, and is especially useful for identifying potential variant sequences (e.g. homologs) that are functional in modulating signal transduction and/or ligand binding. The purpose of screening such combinatorial libraries is to generate, for example, novel *smoothened* homologs which can act as either agonists or antagonist, or alternatively, possess novel activities all together. To illustrate, *smoothened* homologs can be engineered by the present method to provide selective, constitutive activation of *hedgehog* activity, or alternatively, to be dominant negative inhibitors of *smoothened*-dependent signal transduction. For instance, mutagenesis can provide *smoothened* homologs which are able to bind or signal through intracellular regulatory proteins.

In one aspect of this method, the amino acid sequences for a population of *smoothened* homologs from different species or other related proteins are aligned, preferably to promote the highest homology possible. Such a population of variants can include, for example, *smoothened* homologs from one or more species. Amino acids which appear at each position of the aligned sequences are selected to create a degenerate set of combinatorial sequences. In a preferred embodiment, the variegated library of *smoothened* variants is generated by combinatorial mutagenesis at the nucleic acid level, and is encoded by a variegated gene library. For instance, a mixture of synthetic oligonucleotides can be enzymatically ligated into gene sequences such that the degenerate set of potential *smoothened* sequences are expressible as individual polypeptides, or as a library.

There are many ways by which such libraries of potential *smoothened* homologs can be generated from a degenerate oligonucleotide sequence. Chemical synthesis of a degenerate gene sequence can be carried out in an automatic DNA synthesizer, and the synthetic genes then ligated into an appropriate expression vector. The purpose of a degenerate set of genes is to provide, in one mixture, all of the sequences encoding the desired set of potential *smoothened* sequences. The synthesis of degenerate oligonucleotides is well known in the art (see for example, Narang, SA (1983) Tetrahedron 39:3; Itakura et al. (1981) Recombinant DNA, Proc 3rd Cleveland Sympos. Macromolecules, ed. AG Walton, Amsterdam: Elsevier pp273-289; Itakura et al. (1984) Annu. Rev. Biochem. 53:323; Itakura et al. (1984) Science 198:1056; Ike et al. (1983) Nucleic Acid Res. 11:477. Such techniques have been employed in the directed evolution of other proteins (see, for example, Scott et al. (1990) Science 249:386-390; Roberts et al. (1992) PNAS 89:2429-2433; Devlin et al. (1990) Science 249: 404-406; Cwirla et al.

(1990) PNAS 87: 6378-6382; as well as U.S. Patents Nos. 5,223,409, 5,198,346, and 5,096,815).

Likewise, a library of coding sequence fragments can be provided for a *smoothened* clone in order to generate a variegated population of *smoothened* fragments for screening and subsequent selection of bioactive fragments. A variety of techniques are known in the art for generating such libraries, including chemical synthesis. In one embodiment, a library of coding sequence fragments can be generated by (i) treating a double stranded PCR fragment of a *smoothened* coding sequence with a nuclease under conditions wherein nicking occurs only about once per molecule; (ii) denaturing the double stranded DNA; (iii) renaturing the DNA to form double stranded DNA which can include sense/antisense pairs from different nicked products; (iv) removing single stranded portions from reformed duplexes by treatment with S1 nuclease; and (v) ligating the resulting fragment library into an expression vector. By this exemplary method, an expression library can be derived which codes for N-terminal, C-terminal and internal fragments of various sizes.

A wide range of techniques are known in the art for screening gene products of combinatorial libraries made by point mutations or truncation, and for screening cDNA libraries for gene products having a certain property. Such techniques will be generally adaptable for rapid screening of the gene libraries generated by the combinatorial mutagenesis of *smoothened* homologs. The most widely used techniques for screening large gene libraries typically comprises cloning the gene library into replicable expression vectors, transforming appropriate cells with the resulting library of vectors, and expressing the combinatorial genes under conditions in which detection of a desired activity facilitates relatively easy isolation of the vector encoding the gene whose product was detected.

The invention also provides for reduction of the *smoothened* protein to generate mimetics, e.g. peptide or non-peptide agents, which are able to disrupt a biological activity of a wild-type *smoothened* protein, e.g. as inhibitors of protein-protein interactions, such as with *patched*. Thus, such mutagenic techniques as described above are also useful to map the determinants of the *smoothened* proteins which participate in protein-protein interactions. Alternatively, a similar system can be used to derive fragments of a *patched* protein which bind to a *smoothened* protein and competitively inhibit binding of the full length *patched* protein.

To further illustrate, the critical residues of either a *smoothened* protein or a *patched* protein which are involved in molecular recognition of the other can be determined and used to generate *smoothened*-derived or *patched*-derived peptidomimetics which competitively inhibit *patched/smoothened* protein interactions. By employing, for example, scanning mutagenesis to map the amino acid residues of a protein which is involved in binding other proteins, peptidomimetic compounds can be generated which mimic those

residues which facilitate the interaction. Such mimetics may then be used to interfere with the normal function of a *smoothened* protein. For instance, non-hydrolyzable peptide analogs of such residues can be generated using benzodiazepine (e.g., see Freidinger et al. in Peptides: Chemistry and Biology, G.R. Marshall ed., ESCOM Publisher: Leiden, Netherlands, 1988), azepine (e.g., see Huffman et al. in Peptides: Chemistry and Biology, G.R. Marshall ed., ESCOM Publisher: Leiden, Netherlands, 1988), substituted gamma lactam rings (Garvey et al. in Peptides: Chemistry and Biology, G.R. Marshall ed., ESCOM Publisher: Leiden, Netherlands, 1988), keto-methylene pseudopeptides (Ewenson et al. (1986) *J Med Chem* 29:295; and Ewenson et al. in Peptides: Structure and Function (Proceedings of the 9th American Peptide Symposium) Pierce Chemical Co. Rockland, IL, 1985), b-turn dipeptide cores (Nagai et al. (1985) *Tetrahedron Lett* 26:647; and Sato et al. (1986) *J Chem Soc* Perkin Trans 1:1231), and β-aminoalcohols (Gordon et al. (1985) *Biochem Biophys Res Commun* 126:419; and Dann et al. (1986) *Biochem Biophys Res Commun* 134:71).

Another aspect of the invention pertains to an antibody specifically reactive with a *smoothened* protein. For example, by using immunogens derived from a *smoothened* protein, e.g. based on the cDNA sequences, anti-protein/anti-peptide antisera or monoclonal antibodies can be made by standard protocols (See, for example, Antibodies: A Laboratory Manual ed. by Harlow and Lane (Cold Spring Harbor Press: 1988)). A mammal, such as a mouse, a hamster or rabbit can be immunized with an immunogenic form of the peptide (e.g., a *smoothened* polypeptide or an antigenic fragment which is capable of eliciting an antibody response). Techniques for conferring immunogenicity on a protein or peptide include conjugation to carriers or other techniques well known in the art. An immunogenic portion of a *smoothened* protein can be administered in the presence of adjuvant. The progress of immunization can be monitored by detection of antibody titers in plasma or serum. Standard ELISA or other immunoassays can be used with the immunogen as antigen to assess the levels of antibodies. In a preferred embodiment, the subject antibodies are immunospecific for antigenic determinants of a *smoothened* protein of a organism, such as a mammal, e.g. antigenic determinants of a protein represented by SEQ ID No: 5, SEQ ID No: 6, SEQ ID No: 7 and SEQ ID No: 8 or closely related homologs (e.g. at least 70% identical, preferably at least 80% identical, and more preferably at least 90% identical). In yet a further preferred embodiment of the present invention, in order to provide, for example, antibodies which are immuno-selective for discrete *smoothened* homologs, the anti-*smoothened* polypeptide antibodies do not substantially cross react (i.e. does not react specifically) with a protein which is, for example, less than 85%, 90% or 95% homologous with the selected *smoothened*. By "not substantially cross react", it is meant that the antibody has a binding affinity for a non-homologous protein which is at least one order of magnitude, more preferably at least 2 orders of magnitude, and even more preferably at least

3 orders of magnitude less than the binding affinity of the antibody for the intended target *smoothened.*

Following immunization of an animal with an antigenic preparation of a *smoothened* polypeptide, anti-*smoothened* antisera can be obtained and, if desired, polyclonal anti-*smoothened* antibodies isolated from the serum. To produce monoclonal antibodies, antibody-producing cells (lymphocytes) can be harvested from an immunized animal and fused by standard somatic cell fusion procedures with immortalizing cells such as myeloma cells to yield hybridoma cells. Such techniques are well known in the art, an include, for example, the hybridoma technique (originally developed by Kohler and Milstein, (1975) Nature, 256: 495-497), the human B cell hybridoma technique (Kozbar et al., (1983) Immunology Today, 4: 72), and the EBV-hybridoma technique to produce human monoclonal antibodies (Cole et al., (1985) Monoclonal Antibodies and Cancer Therapy, Alan R. Liss, Inc. pp. 77-96). Hybridoma cells can be screened immunochemically for production of antibodies specifically reactive with a *smoothened* polypeptide of the present invention and monoclonal antibodies isolated from a culture comprising such hybridoma cells.

The term antibody as used herein is intended to include fragments thereof which are also specifically reactive with a *smoothened* polypeptide. Antibodies can be fragmented using conventional techniques and the fragments screened for utility in the same manner as described above for whole antibodies. For example, $F(ab)_2$ fragments can be generated by treating antibody with pepsin. The resulting $F(ab)_2$ fragment can be treated to reduce disulfide bridges to produce Fab fragments. The antibody of the present invention is further intended to include bispecific and chimeric molecules having affinity for a *smoothened* protein conferred by at least one CDR region of the antibody.

Both monoclonal and polyclonal antibodies (Ab) directed against authentic *smoothened* polypeptides, or *smoothened* variants, and antibody fragments such as Fab, $F(ab)_2$, Fv and scFv can be used to block the action of a *smoothened* protein and allow the study of the role of these proteins in, for example, differentiation of tissue. Experiments of this nature can aid in deciphering the role of *smoothened* proteins that may be involved in control of proliferation versus differentiation, e.g., in patterning and tissue formation.

Antibodies which specifically bind *smoothened* epitopes can also be used in immunohistochemical staining of tissue samples in order to evaluate the abundance and pattern of expression of each of the subject *smoothened* polypeptides. Anti-*smoothened* antibodies can be used diagnostically in immuno-precipitation and immuno-blotting to detect and evaluate *smoothened* protein levels in tissue as part of a clinical testing procedure. For instance, such measurements can be useful in predictive valuations of the onset or progression of proliferative or differentiative disorders. Likewise, the ability to

monitor *smoothened* protein levels in an individual can allow determination of the efficacy of a given treatment regimen for an individual afflicted with such a disorder. The level of *smoothened* polypeptides may be measured from cells in bodily fluid, such as in samples of cerebral spinal fluid or amniotic fluid, or can be measured in tissue, such as produced by biopsy. Diagnostic assays using anti-*smoothened* antibodies can include, for example, immunoassays designed to aid in early diagnosis of a disorder, particularly ones which are manifest at birth. Diagnostic assays using anti-*smoothened* polypeptide antibodies can also include immunoassays designed to aid in early diagnosis and phenotyping neoplastic or hyperplastic disorders.

Another application of anti-*smoothened* antibodies of the present invention is in the immunological screening of cDNA libraries constructed in expression vectors such as λ gt11, λgt18-23, λZAP, and λORF8. Messenger libraries of this type, having coding sequences inserted in the correct reading frame and orientation, can produce fusion proteins. For instance, λgt11 will produce fusion proteins whose amino termini consist of ß-galactosidase amino acid sequences and whose carboxy termini consist of a foreign polypeptide. Antigenic epitopes of a *smoothened* protein, e.g. orthologs of the *smoothened* protein from other species, can then be detected with antibodies, as, for example, reacting nitrocellulose filters lifted from infected plates with anti-*smoothened* antibodies. Positive phage detected by this assay can then be isolated from the infected plate. Thus, the presence of *smoothened* homologs can be detected and cloned from other animals, as can alternate isoforms (including splicing variants) from humans.

Moreover, the nucleotide sequences determined from the cloning of *smoothened* genes from organisms will further allow for the generation of probes and primers designed for use in identifying and/or cloning *smoothened* homologs in other cell types, e.g. from other tissues, as well as *smoothened* homologs from other organisms. For instance, the present invention also provides a probe/primer comprising a substantially purified oligonucleotide, which oligonucleotide comprises a region of nucleotide sequence that hybridizes under stringent conditions to at least 15, 20, 25 or 30 consecutive nucleotides of sense or anti-sense sequence selected from the group consisting of SEQ ID No: 1, SEQ ID No: 2, SEQ ID No: 3 or SEQ ID No: 4 or naturally occurring mutants thereof. For instance, primers based on the nucleic acid represented in SEQ ID No: 1, SEQ ID No: 2, SEQ ID No: 3 or SEQ ID No: 4, can be used in PCR reactions to clone *smoothened* homologs. Likewise, probes based on the subject *smoothened* sequences can be used to detect transcripts or genomic sequences encoding the same or homologous proteins. In preferred embodiments, the probe further comprises a label group attached thereto and able to be detected, e.g. the label group is selected from amongst radioisotopes, fluorescent compounds, enzymes, and enzyme co-factors.

Such probes can also be used as a part of a diagnostic test kit for identifying cells or tissue which misexpress a *smoothened* protein, such as by measuring a level of a *smoothened*-encoding nucleic acid in a sample of cells from a patient-animal; e.g. detecting *smoothened* mRNA levels or determining whether a genomic *smoothened* gene has been mutated or deleted.

To illustrate, nucleotide probes can be generated from the subject *smoothened* genes which facilitate histological screening of intact tissue and tissue samples for the presence (or absence) of *smoothened*-encoding transcripts. Similar to the diagnostic uses of anti-*smoothened* antibodies, the use of probes directed to *smoothened* messages, or to genomic *smoothened* sequences, can be used for both predictive and therapeutic evaluation of allelic mutations which might be manifest in, for example, degenerative disorders marked by loss of particular cell-types, apoptosis, neoplastic and/or hyperplastic disorders (e.g. unwanted cell growth) or abnormal differentiation of tissue. Used in conjunction with immunoassays as described above, the oligonucleotide probes can help facilitate the determination of the molecular basis for a developmental disorder which may involve some abnormality associated with expression (or lack thereof) of a *smoothened* protein. For instance, variation in polypeptide synthesis can be differentiated from a mutation in a coding sequence.

Accordingly, the present method provides a method for determining if a subject is at risk for a disorder characterized by aberrant apoptosis, cell proliferation and/or differentiation. In preferred embodiments, method can be generally characterized as comprising detecting, in a sample of cells from the subject, the presence or absence of a genetic lesion characterized by at least one of (i) an alteration affecting the integrity of a gene encoding a *smoothened*-protein, or (ii) the mis-expression of the *smoothened* gene. To illustrate, such genetic lesions can be detected by ascertaining the existence of at least one of (i) a deletion of one or more nucleotides from a *smoothened* gene, (ii) an addition of one or more nucleotides to a *smoothened* gene, (iii) a substitution of one or more nucleotides of a *smoothened* gene, (iv) a gross chromosomal rearrangement of a *smoothened* gene, (v) a gross alteration in the level of a messenger RNA transcript of a *smoothened* gene, (vii) aberrant modification of a *smoothened* gene, such as of the methylation pattern of the genomic DNA, (vii) the presence of a non-wild type splicing pattern of a messenger RNA transcript of a *smoothened* gene, (viii) a non-wild type level of a *smoothened*-protein, and (ix) inappropriate post-translational modification of a *smoothened*-protein. As set out below, the present invention provides a large number of assay techniques for detecting lesions in a *smoothened* gene, and importantly, provides the ability to discern between different molecular causes underlying *smoothened*-dependent aberrant cell growth, proliferation and/or differentiation.

In an exemplary embodiment, there is provided a nucleic acid composition comprising a (purified) oligonucleotide probe including a region of nucleotide sequence which is capable of hybridizing to a sense or antisense sequence of a *smoothened* gene, such as represented by any one of SEQ ID Nos: 1-4 and 9-14, or naturally occurring mutants thereof, or 5' or 3' flanking sequences or intronic sequences naturally associated with the subject *smoothened* genes or naturally occurring mutants thereof. The nucleic acid of a cell is rendered accessible for hybridization, the probe is exposed to nucleic acid of the sample, and the hybridization of the probe to the sample nucleic acid is detected. Such techniques can be used to detect lesions at either the genomic or mRNA level, including deletions, substitutions, etc., as well as to determine mRNA transcript levels.

In certain embodiments, detection of the lesion comprises utilizing the probe/primer in a polymerase chain reaction (PCR) (see, e.g. U.S. Patent Nos. 4,683,195 and 4,683,202), such as anchor PCR or RACE PCR, or, alternatively, in a ligation chain reaction (LCR) (see, e.g., Landegran et al. (1988) Science 241:1077-1080; and Nakazawa et al. (1944) PNAS 91:360-364), the later of which can be particularly useful for detecting point mutations in the *smoothened* gene. In a merely illustrative embodiment, the method includes the steps of (i) collecting a sample of cells from a patient, (ii) isolating nucleic acid (e.g., genomic, mRNA or both) from the cells of the sample, (iii) contacting the nucleic acid sample with one or more primers which specifically hybridize to a *smoothened* gene under conditions such that hybridization and amplification of the *smoothened* gene (if present) occurs, and (iv) detecting the presence or absence of an amplification product, or detecting the size of the amplification product and comparing the length to a control sample.

In still another embodiment, the level of a *smoothened*-protein can be detected by immunoassay. For instance, the cells of a biopsy sample can be dispersed, and the level of a *smoothened*-protein present on the surface of each cell can be quantitated by standard immunoassay techniques. In yet another exemplary embodiment, aberrant methylation patterns of a *smoothened* gene can be detected by digesting genomic DNA from a patient sample with one or more restriction endonucleases that are sensitive to methylation and for which recognition sites exist in the *smoothened* gene (including in the flanking and intronic sequences). See, for example, Buiting et al. (1994) Human Mol Genet 3:893-895. Digested DNA is separated by gel electrophoresis, and hybridized with probes derived from, for example, genomic or cDNA sequences. The methylation status of the *smoothened* gene can be determined by comparison of the restriction pattern generated from the sample DNA with that for a standard of known methylation.

A number of techniques exist in the art for identifying ligands to the *smoothened* receptor. For instance, expression cloning can be carried out on a cDNA or genomic library by isolating cells which are decorated with a labeled form of the receptor, such as present in

labeled liposomal preparations. In a preferred embodiment, the technique uses the *smoothened* receptor in an *in situ* assay for detecting *smoothened* ligands in cDNA cloned from tissue samples and whole organisms. For instance, the present invention makes use of the RAP-*in situ* assay (for Receptor Affinity Probe) of Flanagan and Leder (see PCT publications WO 92/06220; and also Cheng et al. (1994) *Cell* 79:157-168). This system involves the use of an expression cloning system whereby a *smoothened* ligand can be cloned on the basis of a cDNA/alkaline phosphatase fusion protein binding to *smoothened.* In general, the method comprises (i) providing a hybrid molecule (the affinity probe) including the cDNA-encoded protein covalently bonded to an enzymatically active tag, preferably for which chromogenic substrates exist, (ii) contacting a cell expressing *smoothened* to form complexes between the probe and *smoothened*, removing unbound probe, and (iii) detecting the affinity complex using a chromogenic substrate for the enzymatic activity associated with the affinity probe.

Furthermore, by making available purified and recombinant *smoothened* polypeptides, the present invention facilitates the development of assays which can be used to screen for drugs which are either agonists or antagonists of the normal cellular function of the subject *smoothened* proteins, or of their role in the pathogenesis of cellular maintenance, differentiation and/or proliferation and disorders related thereto. In a general sense, the assay evaluates the ability of a compound to modulate binding between a *smoothened* protein and a molecule, e.g., *patched* or a *smoothened* ligand, that interacts with the *smoothened* protein. Exemplary compounds which can be screened against such *smoothened*-mediated interactions include peptides, nucleic acids, carbohydrates, small organic molecules, and natural product extract libraries, such as isolated from animals, plants, fungus and/or microbes.

In many drug screening programs which test libraries of compounds and natural extracts, high throughput assays are desirable in order to maximize the number of compounds surveyed in a given period of time. Assays which are performed in cell-free systems, such as may be derived with purified or semi-purified proteins, are often preferred as "primary" screens in that they can be generated to permit rapid development and relatively easy detection of an alteration in a molecular target which is mediated by a test compound. Moreover, the effects of cellular toxicity and/or bioavailability of the test compound can be generally ignored in the *in vitro* system, the assay instead being focused primarily on the effect of the drug on the molecular target as may be manifest in an alteration of binding affinity with a ligand. Accordingly, in an exemplary screening assay of the present invention, a reaction mixture is generated to include a *smoothened* polypeptide, compound(s) of interest, and a "target molecule", e.g., a protein, which interacts with the *smoothened* polypeptide. As set out above, exemplary target molecules include *patched*, extracellular ligands, as well as other protein and non-protein interacting

molecules which interact with *smoothened* in the membrane or cytoplasm. Detection and quantification of interaction of the *smoothened* protein with the target molecule provides a means for determining a compound's efficacy at inhibiting (or potentiating) interaction between the *smoothened* and the target molecule. The efficacy of the compound can be assessed by generating dose response curves from data obtained using various concentrations of the test compound. Moreover, a control assay can also be performed to provide a baseline for comparison. In the control assay, interaction of the *smoothened* polypeptide and target molecule is quantitated in the absence of the test compound.

Interaction between the *smoothened* polypeptide and the target molecule may be detected by a variety of techniques. Modulation of the formation of complexes can be quantitated using, for example, detectably labeled proteins such as radiolabeled, fluorescently labeled, or enzymatically labeled *smoothened* polypeptides, by immunoassay, by chromatographic detection, or by detecting the intrinsic activity of the acetylase.

Accordingly, in an exemplary screening assay for *smoothened* therapeutics, the compound of interest is contacted with a mixture including a *smoothened* protein (e.g., a cell expressing *smoothened*) and a target molecule under conditions in which the two molecules are ordinarily capable of binding one another. To the mixture is then added a composition containing a test compound. Detection and quantification of *smoothened* complexes provides a means for determining the test compound's efficacy at inhibiting (or potentiating) complex formation between the molecules. Moreover, a control assay can also be performed to provide a baseline for comparison. In the control assay, formation of the complexes is quantitated in the absence of the test compound.

In an illustrative embodiment, the screening assay includes all or a suitable portion of the *smoothened* protein, which can be obtained from vertebrate or invertebrate sources. The *smoothened* protein can be provided in the screening assay as a whole protein (preferably expressed on the surface of a cell), or alternatively as a fragment of the full length protein which binds to a target molecule, e.g., an extracellular or intracellular domain(s) as appropriate. In other embodiments, the protein can be provided as part of a liposomal preparation or expressed on the surface of a cell. The *smoothened* protein can derived from a recombinant gene, e.g., being ectopically expressed in a heterologous cell. For instance, the protein can be expressed on oocytes, mammalian cells (e.g., COS, CHO, 3T3 or the like), or yeast cell by standard recombinant DNA techniques. These recombinant cells can be used for receptor binding, signal transduction or gene expression assays. Marigo et al. (1996) *Development* 122:1225-1233 illustrates a binding assay of human *hedgehog* to chick *patched* protein ectopically expressed in *Xenopus laevis* oocytes. The assay system of Marigo et al. can be adapted to the present drug screening assays by

ectopic or endogenous expression of a *smoothened* protein. In the illustrated assay, the amount of *smoothened/patched* complexes can be quantitated by immunoassay or the like.

Complex formation between the *smoothened* polypeptide and a target molecule may be detected by a variety of techniques. For instance, modulation of the formation of complexes can be quantitated using, for example, detectably labeled proteins such as radiolabelled, fluorescently labeled, or enzymatically labeled *hedgehog* polypeptides, by immunoassay, or by chromatographic detection.

Typically, for cell-free assays which utilize an extracellular or intracellular fragment of *smoothened*, it will be desirable to immobilize either the *smoothened* polypeptide or the target polypeptide to facilitate separation of complexes from uncomplexed forms of one of the proteins, as well as to accommodate automation of the assay. In one embodiment, a fusion protein can be provided which adds a domain that allows the protein to be bound to a matrix. For example, glutathione-S-transferase/receptor (GST/receptor) fusion proteins can be adsorbed onto glutathione sepharose beads (Sigma Chemical, St. Louis, MO) or glutathione derivatized microtitre plates, which are then combined with the target polypeptide, e.g. an labeled polypeptide, and the test compound and incubated under conditions conducive to complex formation, e.g. at physiological conditions for salt and pH, though slightly more stringent conditions may be desired. Following incubation, the beads are washed to remove any unbound target polypeptide, and the matrix bead-bound radiolabel determined directly (e.g. beads placed in scintillant), or in the supernatant after the complexes are dissociated. Alternatively, the complexes can be dissociated from the bead, separated by SDS-PAGE gel, and the level of target polypeptide found in the bead fraction quantitated from the gel using standard electrophoretic techniques.

Other techniques for immobilizing proteins on matrices are also available for use in the subject assay. For instance, soluble portions of the *smoothened* protein can be immobilized utilizing conjugation of biotin and streptavidin. For instance, biotinylated receptor molecules can be prepared from biotin-NHS (N-hydroxy-succinimide) using techniques well known in the art (e.g., biotinylation kit, Pierce Chemicals, Rockford, IL), and immobilized in the wells of streptavidin-coated 96 well plates (Pierce Chemical). Alternatively, antibodies reactive with the *smoothened* but which do not interfere with, e.g., *patched* binding, can be derivatized to the wells of the plate and the receptor trapped in the wells by antibody conjugation. As above, preparations of a *smoothened* polypeptide and a test compound are incubated in the wells of the plate, and the amount of complex trapped in the well can be quantitated. Exemplary methods for detecting such complexes, in addition to those described above for the GST-immobilized complexes, include immunodetection of complexes using antibodies reactive with the *smoothened* or target polypeptide; as well as enzyme-linked assays which rely on detecting an enzymatic activity associated with the

smoothened or target polypeptide. In the instance of the latter, the enzyme can be chemically conjugated or provided as a fusion protein, e.g., fused with alkaline phosphatase, and the amount of fusion protein in the complex can be assessed with a chromogenic substrate of the enzyme, e.g. paranitrophenylphosphate. Likewise, a fusion protein including glutathione-S-transferase can be provided, and complex formation quantitated by detecting the GST activity using 1-chloro-2,4-dinitrobenzene (Habig et al (1974) *J Biol Chem* 249:7130).

For processes which rely on immunodetection for quantitating one of the proteins trapped in the complex, antibodies against the protein, such as the anti-*smoothened* antibodies described herein, can be used. Alternatively, the protein to be detected in the complex can be "epitope tagged" in the form of a fusion protein which includes, in addition to the *smoothened* polypeptide, a second polypeptide for which antibodies are readily available (e.g. from commercial sources). For instance, the GST fusion proteins described above can also be used for quantification of binding using antibodies against the GST moiety. Other useful epitope tags include myc-epitopes (e.g., see Ellison et al. (1991) *J Biol Chem* 266:21150-21157) which includes a 10-residue sequence from c-myc, as well as the pFLAG system (International Biotechnologies, Inc.) or the pEZZ-protein A system (Pharamacia, NJ).

Where the desired portion of the *smoothened* protein cannot be provided in soluble form, liposomal vesicles can be used to provide manipulatable and isolatable sources of the protein. For example, both authentic and recombinant forms of the *smoothened* protein can be reconstituted in artificial lipid vesicles (e.g. phosphatidylcholine liposomes) or in cell membrane-derived vesicles (see, for example, Bear et al. (1992) *Cell* 68:809-818; Newton et al. (1983) *Biochemistry* 22:6110-6117; and Reber et al. (1987) *J Biol Chem* 262:11369-11374). Thus, in addition to cell-free assays, such as described above, the readily available source of *smoothened* genes provided by the subject invention also facilitates the generation of cell-based assays for identifying small molecule agonists and antagonists of *smoothened* activity, e.g., which can be used to mimic or inhibit the effect of *hedgehog*. In one embodiment, the interaction of *smoothened* and *patched* proteins in a cell or liposome is assessed.

In addition to characterizing cells that naturally express the *smoothened* protein, cells which have been genetically engineered to ectopically express *smoothened* can be utilized for drug screening assays. As an example, cells which either express low levels or lack expression of the *smoothened* protein, e.g. *Xenopus laevis* oocytes, COS cells or yeast cells, can be genetically modified using standard techniques to ectopically express the *smoothened* protein. (see Marigo et al., *supra*, for analogous expression of *patched*).

The resulting recombinant cells, e.g., which express a functional *smoothened*, can be utilized in receptor binding assays to identify agonist or antagonists its *patched* binding. Binding assays can be performed using whole cells. Furthermore, the recombinant cells of the present invention can be engineered to include other heterologous genes encoding proteins involved in *smoothened*-dependent signal pathways. For example, the gene products of one or more of *costal-2* and/or *fused* can be co-expressed with *smoothened* in the reagent cell, with assays being sensitive to the functional reconstitution of the *smoothened* signal transduction cascade.

Alternatively, liposomal preparations using reconstituted *smoothened* protein can be utilized. *Smoothened* protein purified from detergent extracts from both authentic and recombinant origins can be reconstituted in artificial lipid vesicles (e.g. phosphatidylcholine liposomes) or in cell membrane-derived vesicles (see, for example, Bear et al. (1992) *Cell* 68:809-818; Newton et al. (1983) *Biochemistry* 22:6110-6117; and Reber et al. (1987) *J Biol Chem* 262:11369-11374). The lamellar structure and size of the resulting liposomes can be characterized using electron microscopy. External orientation of the *smoothened* protein in the reconstituted membranes can be demonstrated, for example, by immunoelectron microscopy.

In addition to binding studies, functional assays can be used to identified modulators, i.e., agonists of *smoothened* activities. By detecting changes in intracellular signals, such as alterations in second messengers or gene expression in *smoothened*-expressing cells contacted with a test agent, candidate antagonists and antagonists of *smoothened* signaling can be identified (e.g., having a *hedgehog*-like or *hedgehog*-inhibitory activity).

A number of gene products have been implicated in *smoothened*-mediated signal transduction, including *patched*, the transcription factor *cubitus interruptus* ("Ci" or "Gli" in vertebrates), the serine/threonine kinase *fused* (fu) and the gene products of *costal-2*, and *suppressor of fused*.

The interaction of *smoothened* with *patched*, or the disruption thereof, sets in motion a cascade involving the activation and inhibition of downstream effectors, the ultimate consequence of which is, in some instances, a detectable change in the transcription or translation of a gene. Potential transcriptional targets of *smoothened* signaling are the *patched* gene (Hidalgo and Ingham, 1990 *Development* 110, 291-301; Marigo et al., 1996) and the vertebrate homologs of the drosophila cubitus interruptus gene, the *GLI* genes (Hui et al. (1994) *Dev Biol* 162:402-413). *Patched* gene expression has been shown to be induced in cells of the limb bud and the neural plate that are responsive to *Shh*. (Marigo et al. (1996) *PNAS*, in press; Marigo et al. (1996) *Development* 122:1225-1233). The *GLI* genes encode putative transcription factors having zinc finger DNA

binding domains (Orenic et al. (1990) *Genes & Dev* 4:1053-1067; Kinzler et al. (1990) *Mol Cell Biol* 10:634-642). Transcription of the *GLI* gene has been reported to be upregulated in response to *hedgehog* in limb buds, while transcription of the *GLI3* gene is downregulated in response to *hedgehog* induction (Marigo et al. (1996) *Development* 122:1225-1233). By selecting transcriptional regulatory sequences from such target genes, e.g. from *patched* or *GLI* genes, that are responsible for the up- or down regulation of these genes in response to *smoothened* signaling, and operatively linking such promoters to a reporter gene, one can derive a transcription based assay which is sensitive to the ability of a specific test compound to modify *smoothened* signaling pathways. Expression of the reporter gene, thus, provides a valuable screening tool for the development of compounds that act as antagonists or antagonist of *hedgehog*, e.g., which may be useful as neuroprotective agents and the like.

Reporter gene based assays of this invention measure the end stage of the above described cascade of events, e.g., transcriptional modulation. Accordingly, in practicing one embodiment of the assay, a reporter gene construct is inserted into the reagent cell in order to generate a detection signal dependent on *smoothened* signaling. To identify potential regulatory elements responsive to *smoothened* signaling present in the transcriptional regulatory sequence of a target gene, nested deletions of genomic clones of the target gene can be constructed using standard techniques. See, for example, Current Protocols in Molecular Biology, Ausubel, F.M. et al. (eds.) Greene Publishing Associates, (1989); U.S. Patent 5,266,488; Sato et al. (1995) *J Biol Chem* 270:10314-10322; and Kube et al. (1995) *Cytokine* 7:1-7. A nested set of DNA fragments from the gene's 5'-flanking region are placed upstream of a reporter gene, such as the luciferase gene, and assayed for their ability to direct reporter gene expression in *smoothened* expressing cells. Host cells transiently transfected with reporter gene constructs can be scored for the induction of expression of the reporter gene in the presence and absence of *hedgehog* to determine regulatory sequences which are responsive to *smoothened*-dependent signaling.

In practicing one embodiment of the assay, a reporter gene construct is inserted into the reagent cell in order to generate a detection signal dependent on second messengers generated by induction with *hedgehog* protein. Typically, the reporter gene construct will include a reporter gene in operative linkage with one or more transcriptional regulatory elements responsive to the *hedgehog* activity, with the level of expression of the reporter gene providing the *hedgehog*-dependent and *smoothened*- dependent detection signal. The amount of transcription from the reporter gene may be measured using any method known to those of skill in the art to be suitable. For example, mRNA expression from the reporter gene may be detected using RNAse protection or RNA-based PCR, or the protein product of the reporter gene may be identified by a characteristic stain or an intrinsic activity. The amount of expression from the reporter gene is then compared to the amount of expression

in either the same cell in the absence of the test compound (or *hedgehog*) or it may be compared with the amount of transcription in a substantially identical cell that lacks the target receptor protein. Any statistically or otherwise significant difference in the amount of transcription indicates that the test compound has in some manner altered the signal transduction of the *smoothened* protein, e.g., the test compound is a potential ptc therapeutic.

As described in further detail below, in preferred embodiments the gene product of the reporter is detected by an intrinsic activity associated with that product. For instance, the reporter gene may encode a gene product that, by enzymatic activity, gives rise to a detection signal based on color, fluorescence, or luminescence. In other preferred embodiments, the reporter or marker gene provides a selective growth advantage, e.g., the reporter gene may enhance cell viability, relieve a cell nutritional requirement, and/or provide resistance to a drug.

Preferred reporter genes are those that are readily detectable. The reporter gene may also be included in the construct in the form of a fusion gene with a gene that includes desired transcriptional regulatory sequences or exhibits other desirable properties. Examples of reporter genes include, but are not limited to CAT (chloramphenicol acetyl transferase) (Alton and Vapnek (1979), Nature 282: 864-869) luciferase, and other enzyme detection systems, such as beta-galactosidase; firefly luciferase (deWet et al. (1987), Mol. Cell. Biol. 7:725-737); bacterial luciferase (Engebrecht and Silverman (1984), PNAS 1: 4154-4158; Baldwin et al. (1984), Biochemistry 23: 3663-3667); alkaline phosphatase (Toh et al. (1989) Eur. J. Biochem. 182: 231-238, Hall et al. (1983) J. Mol. Appl. Gen. 2: 101), human placental secreted alkaline phosphatase (Cullen and Malim (1992) Methods in Enzymol. 216:362-368).

Transcriptional control elements which may be included in a reporter gene construct include, but are not limited to, promoters, enhancers, and repressor and activator binding sites. Suitable transcriptional regulatory elements may be derived from the transcriptional regulatory regions of genes whose expression is induced after modulation of a *smoothened* signal transduction pathway. The characteristics of preferred genes from which the transcriptional control elements are derived include, but are not limited to, low or undetectable expression in quiescent cells, rapid induction at the transcriptional level within minutes of extracellular simulation, induction that is transient and independent of new protein synthesis, subsequent shut-off of transcription requires new protein synthesis, and mRNAs transcribed from these genes have a short half-life. It is not necessary for all of these properties to be present.

In yet other embodiments, second messenger generation can be measured directly in the detection step, such as mobilization of intracellular calcium, phospholipid metabolism

or adenylate cyclase activity are quantitated, for instance, the products of phospholipid hydrolysis IP_3, DAG or cAMP could be measured For example, recent studies have implicated protein kinase A (PKA) as a possible component of *hedgehog/smoothened* signaling (Hammerschmidt et al. (1996) *Genes & Dev* 10:647). High PKA activity has been shown to antagonize *hedgehog* signaling in these systems. Conversely, inhibitors of PKA will mimic and/or potentiate the action of *hedgehog*. Although it is unclear whether PKA acts directly downstream or in parallel with *hedgehog* signaling, it is possible that *hedgehog* signaling occurs via inhibition of PKA activity. Thus, detection of PKA activity provides a potential readout for the instant assays.

Smoothened may, under certain circumstances, stimulate the activity of phospholipases. Inositol lipids can be extracted and analyzed using standard lipid extraction techniques. Water soluble derivatives of all three inositol lipids (IP_1, IP_2, IP_3) can also be quantitated using radiolabelling techniques or HPLC.

The mobilization of intracellular calcium or the influx of calcium from outside the cell may also be a response to *smoothened*-dependent signaling, or lack there of. Calcium flux in the reagent cell can be measured using standard techniques. The choice of the appropriate calcium indicator, fluorescent, bioluminescent, metallochromic, or Ca^{++}-sensitive microelectrodes depends on the cell type and the magnitude and time constant of the event under study (Borle (1990) *Environ Health Perspect* 84:45-56). As an exemplary method of Ca^{++} detection, cells could be loaded with the Ca^{++}sensitive fluorescent dye fura-2 or indo-1, using standard methods, and any change in Ca^{++} measured using a fluorometer.

In certain embodiments of the assay, it may be desirable to screen for changes in cellular phosphorylation. As an example, the gene *fused* (fu) which encodes a serine/threonine kinase has been identified as a potential downstream target in *smoothened* signaling. (Preat et al., 1990 *Nature* 347, 87-89; Therond et al. 1993, *Mech. Dev.* 44. 65-80). The ability of compounds to modulate serine/threonine kinase activation could be screened using colony immunoblotting (Lyons and Nelson (1984) *Proc. Natl. Acad. Sci. USA* 81:7426-7430) using antibodies against phosphorylated serine or threonine residues. Reagents for performing such assays are commercially available, for example, phosphoserine and phosphothreonine specific antibodies which measure increases in phosphorylation of those residues can be purchased from commercial sources.

After identifying certain test compounds as potential modulators of one or more bioactivities of a *smoothened* protein, the practitioner of the subject assay will continue to test the efficacy and specificity of the selected compounds both *in vitro* and *in vivo*. Whether for subsequent *in vivo* testing, or for administration to an animal as an approved

drug, agents identified in the subject assay can be formulated in pharmaceutical preparations for *in vivo* administration to an animal, preferably a human.

Another aspect of the present invention relates to a method of inducing and/or maintaining a differentiated state, enhancing survival, and/or inhibiting (or alternatively potentiating) proliferation of a cell, by contacting the cells with an agent which modulates *smoothened*-dependent signal transduction pathways. The subject method could be used to generate and/or maintain an array of different tissue both *in vitro* and *in vivo*. A "*smoothened* therapeutic," whether inhibitory or potentiating with respect to modulating the activity of a *smoothened* protein, can be, as appropriate, any of the preparations described above, including isolated *smoothened* polypeptides (including both agonist and antagonist forms), gene therapy constructs, antisense molecules, peptidomimetics, or agents identified in the drug assays provided herein, e.g., which inhibit or potentiate the interactions of *smoothened* and *patched*.

The *smoothened* therapeutic compounds of the present invention are likely to play an important role in the modulation of cellular proliferation and maintenance of, for example, neuronal, testicular, osteogenic or chondrogenic tissues during disease states. It will also be apparent that, by transient use of modulators of *smoothened* activities, *in vivo* reformation of tissue can be accomplished, e.g. in the development and maintenance of organs such as ectodermal patterning, as well as certain mesodermal and endodermal differentiation processes. By controlling the proliferative and differentiative potential for different cells, the subject *smoothened* therapeutics can be used to reform injured tissue, or to improve grafting and morphology of transplanted tissue. For instance, *smoothened* antagonists and agonists can be employed in a differential manner to regulate different stages of organ repair after physical, chemical or pathological insult. The present method is also applicable to cell culture techniques.

To further illustrate this aspect of the invention, *in vitro* neuronal culture systems have proved to be fundamental and indispensable tools for the study of neural development, as well as the identification of neurotrophic factors such as nerve growth factor (NGF), ciliary trophic factors (CNTF), and brain derived neurotrophic factor (BDNF). Once a neuronal cell has become terminally-differentiated it typically will not change to another terminally differentiated cell-type. However, neuronal cells can nevertheless readily lose their differentiated state. This is commonly observed when they are grown in culture from adult tissue, and when they form a blastema during regeneration. The present method provides a means for ensuring an adequately restrictive environment in order to maintain neuronal cells at various stages of differentiation, and can be employed, for instance, in cell cultures designed to test the specific activities of other trophic factors. In such embodiments of the subject method, the cultured cells can be contacted with a *smoothened*

therapeutic, e.g., such as an agent identified in the assays described above which potentiate *smoothened*-dependent *hedgehog* bioactivities, in order to induce neuronal differentiation (e.g. of a stem cell), or to maintain the integrity of a culture of terminally-differentiated neuronal cells by preventing loss of differentiation. Alternatively, an antagonist of *hedgehog* induction, as certain of the *smoothened* homologs of the present invention are expected to be, can be used to prevent differentiation of progenitor cells in culture.

To further illustrate uses of *smoothened* therapeutics which may be either *hedgehog* agonists or antagonists, it is noted that intracerebral grafting has emerged as an additional approach to central nervous system therapies. For example, one approach to repairing damaged brain tissues involves the transplantation of cells from fetal or neonatal animals into the adult brain (Dunnett et al. (1987) *J Exp Biol* 123:265-289; and Freund et al. (1985) *J Neurosci* 5:603-616). Fetal neurons from a variety of brain regions can be successfully incorporated into the adult brain, and such grafts can alleviate behavioral defects. For example, movement disorder induced by lesions of dopaminergic projections to the basal ganglia can be prevented by grafts of embryonic dopaminergic neurons. Complex cognitive functions that are impaired after lesions of the neocortex can also be partially restored by grafts of embryonic cortical cells. The differential use of *hedgehog* agonists and antagonists in the culture can control the timing and type of differentiation accessible by the culture.

In addition to the implantation of cells cultured in the presence of *hedgehog* agonists and antagonists and other *in vitro* uses, yet another aspect of the present invention concerns the therapeutic application of a *smoothened* therapeutics to enhance survival of neurons and other neuronal cells in both the central nervous system and the peripheral nervous system. The ability of *hedgehog* protein to regulate neuronal differentiation during development of the nervous system and also presumably in the adult state indicates that certain of the *hedgehog* proteins, and accordingly *smoothened* therapeutic which modulate *hedgehog* bioactivities, can be reasonably expected to facilitate control of adult neurons with regard to maintenance, functional performance, and aging of normal cells; repair and regeneration processes in chemically or mechanically lesioned cells; and prevention of degeneration and premature death which result from loss of differentiation in certain pathological conditions. In light of this understanding, the present invention specifically contemplates applications of the subject *smoothened* therapeutics to the treatment of (prevention and/or reduction of the severity of) neurological conditions deriving from: (i) acute, subacute, or chronic injury to the nervous system, including traumatic injury, chemical injury, vasal injury and deficits (such as the ischemia resulting from stroke), together with infectious/inflammatory and tumor-induced injury; (ii) aging of the nervous system including Alzheimer's disease; (iii) chronic neurodegenerative diseases of the nervous system, including Parkinson's disease, Huntington's chorea, amylotrophic lateral sclerosis and the like, as well as spinocerebellar

degenerations; and (iv) chronic immunological diseases of the nervous system or affecting the nervous system, including multiple sclerosis.

Many neurological disorders are associated with degeneration of discrete populations of neuronal elements and may be treatable with a therapeutic regimen which includes a *smoothened* therapeutic that acts as a *hedgehog* agonist. For example, Alzheimer's disease is associated with deficits in several neurotransmitter systems, both those that project to the neocortex and those that reside with the cortex. For instance, the nucleus basalis in patients with Alzheimer's disease have been observed to have a profound (75%) loss of neurons compared to age-matched controls. Although Alzheimer's disease is by far the most common form of dementia, several other disorders can produce dementia. Several of these are degenerative diseases characterized by the death of neurons in various parts of the central nervous system, especially the cerebral cortex. However, some forms of dementia are associated with degeneration of the thalamus or the white matter underlying the cerebral cortex. Here, the cognitive dysfunction results from the isolation of cortical areas by the degeneration of efferents and afferents. Huntington's disease involves the degeneration of intrastraital and cortical cholinergic neurons and GABAergic neurons. Pick's disease is a severe neuronal degeneration in the neocortex of the frontal and anterior temporal lobes, sometimes accompanied by death of neurons in the striatum. Treatment of patients suffering from such degenerative conditions can include the application of *smoothened* therapeutics in order to control, for example, differentiation and apoptotic events which give rise to loss of neurons (e.g. to enhance survival of existing neurons) as well as promote differentiation and repopulation by progenitor cells in the area affected.

In addition to degenerative-induced dementias, a pharmaceutical preparation of one or more of the subject *smoothened* therapeutics can be applied opportunely in the treatment of neurodegenerative disorders which have manifestations of tremors and involuntary movements. Parkinson's disease, for example, primarily affects subcortical structures and is characterized by degeneration of the nigrostriatal pathway, raphe nuclei, locus cereleus, and the motor nucleus of vagus. Ballism is typically associated with damage to the subthalmic nucleus, often due to acute vascular accident. Also included are neurogenic and myopathic diseases which ultimately affect the somatic division of the peripheral nervous system and are manifest as neuromuscular disorders. Examples include chronic atrophies such as amyotrophic lateral sclerosis, Guillain-Barre syndrome and chronic peripheral neuropathy, as well as other diseases which can be manifest as progressive bulbar palsies or spinal muscular atrophies. The present method is amenable to the treatment of disorders of the cerebellum which result in hypotonia or ataxia, such as those lesions in the cerebellum which produce disorders in the limbs ipsilateral to the lesion. For instance, a preparation of a *smoothened* therapeutic can used to treat a restricted form of cerebellar cortical

degeneration involving the anterior lobes (vermis and leg areas) such as is common in alcoholic patients.

In an illustrative embodiment, the subject method is used to treat amyotrophic lateral sclerosis. ALS is a name given to a complex of disorders that comprise upper and lower motor neurons. Patients may present with progressive spinal muscular atrophy, progressive bulbar palsy, primary lateral sclerosis, or a combination of these conditions. The major pathological abnormality is characterized by a selective and progressive degeneration of the lower motor neurons in the spinal cord and the upper motor neurons in the cerebral cortex. The therapeutic application of a *smoothened* therapeutic which is a *hedgehog* agonist can be used alone, or in conjunction with other neurotrophic factors such as CNTF, BDNF or NGF to prevent and/or reverse motor neuron degeneration in ALS patients.

Smoothened therapeutics of the present invention can also be used in the treatment of autonomic disorders of the peripheral nervous system, which include disorders affecting the innervation of smooth muscle and endocrine tissue (such as glandular tissue). For instance, the subject method can be used to treat tachycardia or atrial cardiac arrythmias which may arise from a degenerative condition of the nerves innervating the striated muscle of the heart.

Furthermore, a potential role for certain of the *smoothened* therapeutics derives from the role of *hedgehog* proteins in development and maintenance of dendritic processes of axonal neurons. Potential roles for *hedgehog* agonists consequently include guidance for axonal projections and the ability to promote differentiation and/or maintenance of the innervating cells to their axonal processes. Accordingly, compositions comprising *smoothened* therapeutics which agonize *hedgehog* activity, may be employed to support the survival and reprojection of several types of ganglionic neurons sympathetic and sensory neurons as well as motor neurons. In particular, such therapeutic compositions may be useful in treatments designed to rescue, for example, various neurons from lesion-induced death as well as guiding reprojection of these neurons after such damage. Such diseases include, but are not limited to, CNS trauma infarction, infection (such as viral infection with varicella-zoster), metabolic disease, nutritional deficiency, toxic agents (such as cisplatin treatment).

Moreover, certain of the *smoothened* therapeutics (e.g., which antagonize *hedgehog* induction) may be useful in the selective ablation of sensory neurons, for example, in the treatment of chronic pain syndromes.

As appropriate, *smoothened* therapeutics can be used in nerve prostheses for the repair of central and peripheral nerve damage. In particular, where a crushed or severed axon is intubulated by use of a prosthetic device, certain of *smoothened* therapeutics can be added to the prosthetic device to increase the rate of growth and regeneration of the

dendridic processes. Exemplary nerve guidance channels are described in U.S. patents 5,092,871 and 4,955,892. Accordingly, a severed axonal process can be directed toward the nerve ending from which it was severed by a prosthesis nerve guide.

In another embodiment, the subject method can be used in the treatment of neoplastic or hyperplastic transformations such as may occur in the central nervous system. For instance, certain of the *smoothened* therapeutics which induce differentiation of neuronal cells can be utilized to cause such transformed cells to become either post-mitotic or apoptotic. Treatment with a *smoothened* therapeutic may facilitate disruption of autocrine loops, such as TGF-β or PDGF autostimulatory loops, which are believed to be involved in the neoplastic transformation of several neuronal tumors. *smoothened* therapeutics may, therefore, thus be of use in the treatment of, for example, malignant gliomas, medulloblastomas, neuroectodermal tumors, and ependymonas.

Yet another aspect of the present invention concerns the application of the discovery that *hedgehog* proteins are morphogenic signals involved in other vertebrate organogenic pathways in addition to neuronal differentiation as described above, having apparent roles in other endodermal patterning, as well as both mesodermal and endodermal differentiation processes. As described in the literature, *Shh* plays a role in proper limb growth and patterning by initiating expression of signaling molecules, including *Bmp-2* in the mesoderm and *Fgf-4* in the ectoderm. Thus, it is contemplated by the invention that compositions comprising certain of the *smoothened* therapeutics can also be utilized for both cell culture and therapeutic methods involving generation and maintenance of non-neuronal tissue.

In one embodiment, the present invention makes use of the discovery that *hedgehog* proteins, such as *Shh*, are apparently involved in controlling the development of stem cells responsible for formation of the digestive tract, liver, lungs, and other organs which derive from the primitive gut. *Shh* serves as an inductive signal from the endoderm to the mesoderm, which is critical to gut morphogenesis. Therefore, for example, *hedgehog* agonists can be employed in the development and maintenance of an artificial liver which can have multiple metabolic functions of a normal liver. In an exemplary embodiment, a *smoothened* therapeutic which acts as a *hedgehog* agonist can be used to induce differentiation of digestive tube stem cells to form hepatocyte cultures which can be used to populate extracellular matrices, or which can be encapsulated in biocompatible polymers, to form both implantable and extracorporeal artificial livers.

In another embodiment, therapeutic compositions of *hedgehog* agonists can be utilized in conjunction with transplantation of such artificial livers, as well as embryonic liver structures, to promote intraperitoneal implantation, vascularization, and *in vivo* differentiation and maintenance of the engrafted liver tissue.

In yet another embodiment, *smoothened* therapeutics can be employed therapeutically to regulate such organs after physical, chemical or pathological insult. For instance, therapeutic compositions comprising *hedgehog* agonists can be utilized in liver repair subsequent to a partial hepatectomy. Similarly, therapeutic compositions containing *hedgehog* agonists can be used to promote regeneration of lung tissue in the treatment of emphysema.

In still another embodiment of the present invention, compositions comprising *smoothened* therapeutics can be used in the *in vitro* generation of skeletal tissue, such as from skeletogenic stem cells, as well as the *in vivo* treatment of skeletal tissue deficiencies. The present invention particularly contemplates the use of *smoothened* therapeutics which agonize a *hedgehog* a skeletogenic activity, such as an ability to induce chondrogenesis and/or osteogenesis. By "skeletal tissue deficiency", it is meant a deficiency in bone or other skeletal connective tissue at any site where it is desired to restore the bone or connective tissue, no matter how the deficiency originated, e.g. whether as a result of surgical intervention, removal of tumor, ulceration, implant, fracture, or other traumatic or degenerative conditions.

For instance, the present invention makes available effective therapeutic methods and compositions for restoring cartilage function to a connective tissue. Such methods are useful in, for example, the repair of defects or lesions in cartilage tissue which is the result of degenerative wear such as that which results in arthritis, as well as other mechanical derangements which may be caused by trauma to the tissue, such as a displacement of torn meniscus tissue, meniscectomy, a laxation of a joint by a torn ligament, malignment of joints, bone fracture, or by hereditary disease. The present reparative method is also useful for remodeling cartilage matrix, such as in plastic or reconstructive surgery, as well as periodontal surgery. The present method may also be applied to improving a previous reparative procedure, for example, following surgical repair of a meniscus, ligament, or cartilage. Furthermore, it may prevent the onset or exacerbation of degenerative disease if applied early enough after trauma.

In one embodiment of the present invention, the subject method comprises treating the afflicted connective tissue with a therapeutically sufficient amount of a *hedgehog* agonist, particularly *smoothened* therapeutic which agonizes *Ihh* activity, to generate a cartilage repair response in the connective tissue by stimulating the differentiation and/or proliferation of chondrocytes embedded in the tissue. Induction of chondrocytes by treatment with a *hedgehog* agonist can subsequently result in the synthesis of new cartilage matrix by the treated cells. Such connective tissues as articular cartilage, interarticular cartilage (menisci), costal cartilage (connecting the true ribs and the sternum), ligaments, and tendons are particularly amenable to treatment in reconstructive and/or regenerative

therapies using the subject method. As used herein, regenerative therapies include treatment of degenerative states which have progressed to the point of which impairment of the tissue is obviously manifest, as well as preventive treatments of tissue where degeneration is in its earliest stages or imminent. The subject method can further be used to prevent the spread of mineralisation into fibrotic tissue by maintaining a constant production of new cartilage.

In an illustrative embodiment, the subject method can be used to treat cartilage of a diarthroidal joint, such as a knee, an ankle, an elbow, a *smoothened*, a wrist, a knuckle of either a finger or toe, or a temperomandibular joint. The treatment can be directed to the meniscus of the joint, to the articular cartilage of the joint, or both. To further illustrate, the subject method can be used to treat a degenerative disorder of a knee, such as which might be the result of traumatic injury (e.g., a sports injury or excessive wear) or osteoarthritis. An injection of a *smoothened* therapeutic into the joint with, for instance, an arthroscopic needle, can be used to treat the afflicted cartilage. In some instances, the injected agent can be in the form of a hydrogel or other slow release vehicle described above in order to permit a more extended and regular contact of the agent with the treated tissue.

The present invention further contemplates the use of the subject method in the field of cartilage transplantation and prosthetic device therapies. To date, the growth of new cartilage from either transplantation of autologous or allogenic cartilage has been largely unsuccessful. Problems arise, for instance, because the characteristics of cartilage and fibrocartilage varies between different tissue: such as between articular, meniscal cartilage, ligaments, and tendons, between the two ends of the same ligament or tendon, and between the superficial and deep parts of the tissue. The zonal arrangement of these tissues may reflect a gradual change in mechanical properties, and failure occurs when implanted tissue, which has not differentiated under those conditions, lacks the ability to appropriately respond. For instance, when meniscal cartilage is used to repair anterior cruciate ligaments, the tissue undergoes a metaplasia to pure fibrous tissue. By promoting chondrogenesis, the subject method can be used to particularly addresses this problem, by causing the implanted cells to become more adaptive to the new environment and effectively resemble hypertrophic chondrocytes of an earlier developmental stage of the tissue. Thus, the action of chondrogenesis in the implanted tissue, as provided by the subject method, and the mechanical forces on the actively remodeling tissue can synergize to produce an improved implant more suitable for the new function to which it is to be put.

In similar fashion, the subject method can be applied to enhancing both the generation of prosthetic cartilage devices and to their implantation. The need for improved treatment has motivated research aimed at creating new cartilage that is based on collagen-glycosaminoglycan templates (Stone et al. (1990) *Clin Orthop Relat Red* 252:129), isolated

chondrocytes (Grande et al. (1989) *J Orthop Res* 7:208; and Takigawa et al. (1987) *Bone Miner* 2:449), and chondrocytes attached to natural or synthetic polymers (Walitani et al. (1989) *J Bone Jt Surg* 71B:74; Vacanti et al. (1991) *Plast Reconstr Surg* 88:753; von Schroeder et al. (1991) *J Biomed Mater Res* 25:329; Freed et al. (1993) *J Biomed Mater Res* 27:11; and the Vacanti et al. U.S. Patent No. 5,041,138). For example, chondrocytes can be grown in culture on biodegradable, biocompatible highly porous scaffolds formed from polymers such as polyglycolic acid, polylactic acid, agarose gel, or other polymers which degrade over time as function of hydrolysis of the polymer backbone into innocuous monomers. The matrices are designed to allow adequate nutrient and gas exchange to the cells until engraftment occurs. The cells can be cultured *in vitro* until adequate cell volume and density has developed for the cells to be implanted. One advantage of the matrices is that they can be cast or molded into a desired shape on an individual basis, so that the final product closely resembles the patient's own ear or nose (by way of example), or flexible matrices can be used which allow for manipulation at the time of implantation, as in a joint.

In one embodiment of the subject method, the implants are contacted with a *smoothened* therapeutic during the culturing process, such as an *Ihh* agonist, in order to induce and/or maintain differentiated chondrocytes in the culture in order as to further stimulate cartilage matrix production within the implant. In such a manner, the cultured cells can be caused to maintain a phenotype typical of a chondrogenic cell (i.e. hypertrophic), and hence continue the population of the matrix and production of cartilage tissue.

In another embodiment, the implanted device is treated with a *smoothened* therapeutic in order to actively remodel the implanted matrix and to make it more suitable for its intended function. As set out above with respect to tissue transplants, the artificial transplants suffer from the same deficiency of not being derived in a setting which is comparable to the actual mechanical environment in which the matrix is implanted. The activation of the chondrocytes in the matrix by the subject method can allow the implant to acquire characteristics similar to the tissue for which it is intended to replace.

In yet another embodiment, the subject method is used to enhance attachment of prosthetic devices. To illustrate, the subject method can be used in the implantation of a periodontal prosthesis, wherein the treatment of the surrounding connective tissue stimulates formation of periodontal ligament about the prosthesis, as well as inhibits formation of fibrotic tissue proximate the prosthetic device.

In still further embodiments, the subject method can be employed for the generation of bone (osteogenesis) at a site in the animal where such skeletal tissue is deficient. Indian *hedgehog* is particularly associated with the hypertrophic chondrocytes that are ultimately replaced by osteoblasts. For instance, administration of a *smoothened* therapeutic of the

present invention can be employed as part of a method for treating bone loss in a subject, e.g. to prevent and/or reverse osteoporosis and other osteopenic disorders, as well as to regulate bone growth and maturation. For example, preparations comprising *hedgehog* agonists can be employed, for example, to induce endochondral ossification, at least so far as to facilitate the formation of cartilaginous tissue precursors to form the "model" for ossification. Therapeutic compositions of *smoothened* therapeutics can be supplemented, if required, with other osteoinductive factors, such as bone growth factors (e.g. TGF-β factors, such as the bone morphogenetic factors *BMP-2* and *BMP-4*, as well as activin), and may also include, or be administered in combination with, an inhibitor of bone resorption such as estrogen, bisphosphonate, sodium fluoride, calcitonin, or tamoxifen, or related compounds. However, it will be appreciated that *hedgehog* proteins, such as I*hh* and S*hh* are likely to be upstream of BMPs, e.g. treatment with a *hedgehog* agonist will have the advantage of initiating endogenous expression of BMPs along with other factors.

In yet another embodiment, the *smoothened* therapeutic of the present invention can be used in the treatment of testicular cells, so as to modulate spermatogenesis. In light of the finding that *hedgehog* proteins are involved in the differentiation and/or proliferation and maintenance of testicular germ cells, *hedgehog* antagonist can be utilized to block the action of a naturally-occurring *hedgehog* protein. In a preferred embodiment, the *smoothened* therapeutic inhibits the biological activity of *Dhh* with respect to spermatogenesis, by competitively binding *hedgehog* in the testis. That is, the *smoothened* therapeutic can be administered as a contraceptive formulation. Alternatively, *smoothened* therapeutics which agonize the spermatogenic activity of *Dhh* can be used as fertility enhancers. In similar fashion, *hedgehog* agonists and antagonists are potentially useful for modulating normal ovarian function.

The source of the *smoothened* therapeutics to be formulated will depend on the particular form of the agent. Small organic molecules and peptidyl fragments can be chemically synthesized and provided in a pure form suitable for pharmaceutical/cosmetic usage. Products of natural extracts can be purified according to techniques known in the art.

The *smoothened* therapeutic formulations used in the method of the invention are most preferably applied in the form of appropriate compositions. As appropriate compositions there may be cited all compositions usually employed for systemically or locally (such as intrathecal) administering drugs. The pharmaceutically acceptable carrier should be substantially inert, so as not to act with the active component. Suitable inert carriers include water, alcohol polyethylene glycol, mineral oil or petroleum gel, propylene glycol and the like.

To prepare the pharmaceutical compositions of this invention, an effective amount of the particular *smoothened* therapeutic as the active ingredient is combined in intimate

admixture with a pharmaceutically acceptable carrier, which carrier may take a wide variety of forms depending on the form of preparation desired for administration. These pharmaceutical compositions are desirable in unitary dosage form suitable, particularly, for administration orally, rectally, percutaneously, or by parenteral injection. For example, in preparing the compositions in oral dosage form, any of the usual pharmaceutical media may be employed such as, for example, water, glycols, oils, alcohols and the like in the case of oral liquid preparations such as suspensions, syrups, elixirs and solutions; or solid carriers such as starches, sugars, kaolin, lubricants, binders, disintegrating agents and the like in the case of powders, pills, capsules, and tablets. Because of their ease in administration, tablets and capsules represents the most advantageous oral dosage unit form, in which case solid pharmaceutical carriers are obviously employed. For parenteral compositions, the carrier will usually comprise sterile water, at least in large part, though other ingredients, for example, to aid solubility, may be included. Injectable solutions, for example, may be prepared in which the carrier comprises saline solution, glucose solution or a mixture of saline and glucose solution. Injectable suspensions may also be prepared in which case appropriate liquid carriers, suspending agents and the like may be employed. Also included are solid form preparations which are intended to be converted, shortly before use, to liquid form preparations. In the compositions suitable for percutaneous administration, the carrier optionally comprises a penetration enhancing agent and/or a suitable wetting agent, optionally combined with suitable additives of any nature in minor proportions, which additives do not introduce a significant deleterious effect on the skin.

It is especially advantageous to formulate the subject compositions in dosage unit form for ease of administration and uniformity of dosage. Dosage unit form as used in the specification and claims herein refers to physically discrete units suitable as unitary dosages, each unit containing a predetermined quantity of active ingredient calculated to produce the desired therapeutic effect in association with the required pharmaceutical carrier. Examples of such dosage unit forms are tablets (including scored or coated tablets), capsules, pills, powders packets, wafers, injectable solutions or suspensions, teaspoonfuls, tablespoonfuls and the like, and segregated multiples thereof.

The pharmaceutical preparations of the present invention can be used, as stated above, for the many applications which can be considered cosmetic uses. Cosmetic compositions known in the art, preferably hypoallergic and pH controlled are especially preferred, and include toilet waters, packs, lotions, skin milks or milky lotions. The preparations contain, besides the *smoothened* therapeutic, components usually employed in such preparations. Examples of such components are oils, fats, waxes, surfactants, humectants, thickening agents, antioxidants, viscosity stabilizers, chelating agents, buffers, preservatives, perfumes, dyestuffs, lower alkanols, and the like. If desired, further ingredients may be incorporated in the compositions, e.g. anti-inflammatory agents,

antibacterials, antifungals, disinfectants, vitamins, sunscreens, antibiotics, or other anti-acne agents.

Examples of oils comprise fats and oils such as olive oil and hydrogenated oils; waxes such as beeswax and lanolin; hydrocarbons such as liquid paraffin, ceresin, and squalane; fatty acids such as stearic acid and oleic acid; alcohols such as cetyl alcohol, stearyl alcohol, lanolin alcohol, and hexadecanol; and esters such as isopropyl myristate, isopropyl palmitate and butyl stearate. As examples of surfactants there may be cited anionic surfactants such as sodium stearate, sodium cetylsulfate, polyoxyethylene laurylether phosphate, sodium N-acyl glutamate; cationic surfactants such as stearyldimethylbenzylammonium chloride and stearyltrimethylammonium chloride; ampholytic surfactants such as alkylaminoethylglycine hydrochloride solutions and lecithin; and nonionic surfactants such as glycerin monostearate, sorbitan monostearate, sucrose fatty acid esters, propylene glycol monostearate, polyoxyethylene oleylether, polyethylene glycol monostearate, polyoxyethylene sorbitan monopalmitate, polyoxyethylene coconut fatty acid monoethanolamide, polyoxypropylene glycol (e.g. the materials sold under the trademark "Pluronic"), polyoxyethylene castor oil, and polyoxyethylene lanolin. Examples of humectants include glycerin, 1,3-butylene glycol, and propylene glycol; examples of lower alcohols include ethanol and isopropanol; examples of thickening agents include xanthan gum, hydroxypropyl cellulose, hydroxypropyl methyl cellulose, polyethylene glycol and sodium carboxymethyl cellulose; examples of antioxidants comprise butylated hydroxytoluene, butylated hydroxyanisole, propyl gallate, citric acid and ethoxyquin; examples of chelating agents include disodium edetate and ethanehydroxy diphosphate; examples of buffers comprise citric acid, sodium citrate, boric acid, borax, and disodium hydrogen phosphate; and examples of preservatives are methyl parahydroxybenzoate, ethyl parahydroxybenzoate, dehydroacetic acid, salicylic acid and benzoic acid.

For preparing ointments, creams, toilet waters, skin milks, and the like, typically from 0.01 to 10% in particular from 0.1 to 5% and more in particular from 0.2 to 2.5% of the active ingredient, e.g., of the *smoothened* therapeutic, will be incorporated in the compositions. In ointments or creams, the carrier for example consists of 1 to 20%, in particular 5 to 15% of a humectant, 0.1 to 10% in particular from 0.5 to 5% of a thickener and water; or said carrier may consist of 70 to 99%, in particular 20 to 95% of a surfactant, and 0 to 20%, in particular 2.5 to 15% of a fat; or 80 to 99.9% in particular 90 to 99% of a thickener; or 5 to 15% of a surfactant, 2-15% of a humectant, 0 to 80% of an oil, very small (< 2%) amounts of preservative, coloring agent and/or perfume, and water. In a toilet water, the carrier for example consists of 2 to 10% of a lower alcohol, 0.1 to 10% or in particular 0.5 to 1% of a surfactant, 1 to 20%, in particular 3 to 7% of a humectant, 0 to 5% of a buffer, water and small amounts (< 2%) of preservative, dyestuff and/or perfume. In a skin milk, the carrier typically consists of 10-50% of oil, 1 to 10% of surfactant, 50-80% of

water and 0 to 3% of preservative and/or perfume. In the aforementioned preparations, all % symbols refer to weight by weight percentage.

Particular compositions for use in the method of the present invention are those wherein the *smoothened* therapeutic is formulated in liposome-containing compositions. Liposomes are artificial vesicles formed by amphiphatic molecules such as polar lipids, for example, phosphatidyl cholines, ethanolamines and serines, sphingomyelins, cardiolipins, plasmalogens, phosphatidic acids and cerebiosides. Liposomes are formed when suitable amphiphathic molecules are allowed to swell in water or aqueous solutions to form liquid crystals usually of multilayer structure comprised of many bilayers separated from each other by aqueous material (also referred to as coarse liposomes). Another type of liposome known to be consisting of a single bilayer encapsulating aqueous material is referred to as a unilamellar vesicle. If water-soluble materials are included in the aqueous phase during the swelling of the lipids they become entrapped in the aqueous layer between the lipid bilayers.

Water-soluble active ingredients are encapsulated in the aqueous spaces between the molecular layers. The lipid soluble active ingredient of a *smoothened* therapeutic, such as an organic mimetic, is predominantly incorporated into the lipid layers, although polar head groups may protrude from the layer into the aqueous space. The encapsulation of these compounds can be achieved by a number of methods. The method most commonly used involves casting a thin film of phospholipid onto the walls of a flask by evaporation from an organic solvent. When this film is dispersed in a suitable aqueous medium, multilamellar liposomes are formed. Upon suitable sonication, the coarse liposomes form smaller similarly closed vesicles.

Water-soluble active ingredients are usually incorporated by dispersing the cast film with an aqueous solution of the compound. The unencapsulated compound is then removed by centrifugation, chromatography, dialysis or other art-known suitable procedures. The lipid-soluble active ingredient is usually incorporated by dissolving it in the organic solvent with the phospholipid prior to casting the film. If the solubility of the material in the lipid phase is not exceeded or the amount present is not in excess of that which can be bound to the lipid, liposomes prepared by the above method usually contain most of the material bound in the lipid bilayers; separation of the liposomes from unencapsulated material is not required.

A particularly convenient method for preparing liposome formulated forms of *smoothened* therapeutics is the method described in EP-A-253,619, incorporated herein by reference. In this method, single bilayered liposomes containing encapsulated active ingredients are prepared by dissolving the lipid component in an organic medium, injecting the organic solution of the lipid component under pressure into an aqueous component

while simultaneously mixing the organic and aqueous components with a high speed homogenizer or mixing means, whereupon the liposomes are formed spontaneously.

The single bilayered liposomes containing the encapsulated *smoothened* therapeutic can be employed directly or they can be employed in a suitable pharmaceutically acceptable carrier for localized administration. The viscosity of the liposomes can be increased by the addition of one or more suitable thickening agents such as, for example xanthan gum, hydroxypropyl cellulose, hydroxypropyl methylcellulose and mixtures thereof. The aqueous component may consist of water alone or it may contain electrolytes, buffered systems and other ingredients, such as, for example, preservatives. Suitable electrolytes which can be employed include metal salts such as alkali metal and alkaline earth metal salts. The preferred metal salts are calcium chloride, sodium chloride and potassium chloride. The concentration of the electrolyte may vary from zero to 260 mM, preferably from 5 mM to 160 mM. The aqueous component is placed in a suitable vessel which can be adapted to effect homogenization by effecting great turbulence during the injection of the organic component. Homogenization of the two components can be accomplished within the vessel, or, alternatively, the aqueous and organic components may be injected separately into a mixing means which is located outside the vessel. In the latter case, the liposomes are formed in the mixing means and then transferred to another vessel for collection purpose.

The organic component consists of a suitable non-toxic, pharmaceutically acceptable solvent such as, for example ethanol, glycerol, propylene glycol and polyethylene glycol, and a suitable phospholipid which is soluble in the solvent. Suitable phospholipids which can be employed include lecithin, phosphatidylcholine, phosphatydylserine, phosphatidylethanol-amine, phosphatidylinositol, lysophosphatidylcholine and phosphatidyl glycerol, for example. Other lipophilic additives may be employed in order to selectively modify the characteristics of the liposomes. Examples of such other additives include stearylamine, phosphatidic acid, tocopherol, cholesterol and lanolin extracts.

In addition, other ingredients which can prevent oxidation of the phospholipids may be added to the organic component. Examples of such other ingredients include tocopherol, butylated hydroxyanisole, butylated hydroxytoluene, ascorbyl palmitate and ascorbyl oleate. Preservatives such a benzoic acid, methyl paraben and propyl paraben may also be added.

Methods of introduction may also be provided by rechargeable or biodegradable devices. Various slow release polymeric devices have been developed and tested *in vivo* in recent years for the controlled delivery of drugs, including proteinacious biopharmaceuticals. A variety of biocompatible polymers (including hydrogels), including both biodegradable and non-degradable polymers, can be used to form an implant for the sustained release of an *hh* at a particular target site.

An essential feature of certain embodiments of the implant can be the linear release of the therapeutic, which can be achieved through the manipulation of the polymer composition and form. By choice of monomer composition or polymerization technique, the amount of water, porosity and consequent permeability characteristics can be controlled. The selection of the shape, size, polymer, and method for implantation can be determined on an individual basis according to the disorder to be treated and the individual patient response. The generation of such implants is generally known in the art. See, for example, *Concise Encylopedia of Medical & Dental Materials*, ed. by David Williams (MIT Press: Cambridge, MA, 1990); and the Sabel et al. U.S. Patent No. 4,883,666.

In another embodiment of an implant, a source of cells producing the therapeutic, e.g., secreting a soluble form of a *smoothened* ligand, is encapsulated in implantable hollow fibers or the like. Such fibers can be pre-spun and subsequently loaded with the cell source (Aebischer et al. U.S. Patent No. 4,892,538; Aebischer et al. U.S. Patent No. 5,106.627; Hoffman et al. (1990) *Expt. Neurobiol.* 110:39-44; Jaeger et al. (1990) *Prog. Brain Res.* 82:41-46; and Aebischer et al. (1991) *J. Biomech. Eng.* 113:178-183), or can be co-extruded with a polymer which acts to form a polymeric coat about the cells (Lim U.S. Patent No. 4,391,909; Sefton U.S. Patent No. 4,353.888; Sugamori et al. (1989) *Trans. Am. Artif. Intern. Organs* 35:791-799; Sefton et al. (1987) *Biotehnol. Bioeng.* 29:1135-1143; and Aebischer et al. (1991) *Biomaterials* 12:50-55).

Another aspect of the invention features transgenic non-human animals which express a heterologous *smoothened* gene of the present invention, and/or which have had one or more genomic *smoothened* genes disrupted in at least a tissue or cell-types of the animal. Accordingly, the invention features an animal model for developmental diseases, which animal has one or more *smoothened* allele which is mis-expressed. For example, an animal can be generated which has one or more *smoothened* alleles deleted or otherwise rendered inactive. Such a model can then be used to study disorders arising from mis-expressed *smoothened* genes, as well as for evaluating potential therapies for similar disorders.

The transgenic animals of the present invention all include within a plurality of their cells a transgene of the present invention, which transgene alters the phenotype of the "host cell" with respect to regulation by the *smoothened protein*, e.g., of cell growth, death and/or differentiation. Since it is possible to produce transgenic organisms of the invention utilizing one or more of the transgene constructs described herein, a general description will be given of the production of transgenic organisms by referring generally to exogenous genetic material. This general description can be adapted by those skilled in the art in order to incorporate specific transgene sequences into organisms utilizing the methods and materials described herein and those generally known in the art.

In one embodiment, the transgene construct is a knockout construct. Such transgene constructs usually are insertion-type or replacement-type constructs (Hasty et al. (1991) *Mol Cell Biol* 11:4509). The transgene constructs for disruption of a *smoothened* gene are designed to facilitate homologous recombination with a portion of the genomic *smoothened* gene so as to prevent the functional expression of the endogenous *smoothened* gene. In preferred embodiments, the nucleotide sequence used as the knockout construct can be comprised of (1) DNA from some portion of the endogenous *smoothened* gene (exon sequence, intron sequence, promoter sequences, etc.) which direct recombination and (2) a marker sequence which is used to detect the presence of the knockout construct in the cell. The knockout construct is inserted into a cell, and integrates with the genomic DNA of the cell in such a position so as to prevent or interrupt transcription of the native *smoothened* gene. Such insertion can occur by homologous recombination, i.e., regions of the knockout construct that are homologous to the endogenous *smoothened* gene sequence hybridize to the genomic DNA and recombine with the genomic sequences so that the construct is incorporated into the corresponding position of the genomic DNA. The knockout construct can comprise (1) a full or partial sequence of one or more exons and/or introns of the *smoothened* gene to be disrupted, (2) sequences which flank the 5' and 3' ends of the coding sequence of the *smoothened* gene, or (3) a combination thereof.

A preferred knockout construct will delete, by targeted homologous recombination, essential structural elements of an endogenous *smoothened* gene. For example, the targeting construct can recombine with the genomic *smoothened* gene can delete a portion of the coding sequence, and/or essential transcriptional regulatory sequences of the gene.

Alternatively, the knockout construct can be used to interrupt essential structural and/or regulatory elements of an endogenous *smoothened* gene by targeted insertion of a polynucleotide sequence. For instance, a knockout construct can recombine with a *smoothened* gene and insert a nonhomologous sequence, such as a *neo* expression cassette, into a structural element (e.g., an exon) and/or regulatory element (e.g., enhancer, promoter, intron splice site, polyadenylation site, etc.) to yield a targeted *smoothened* allele having an insertional disruption. The inserted nucleic acid can range in size from 1 nucleotide (e.g., to produce a frameshift) to several kilobases or more, and is limited only by the efficiency of the targeting technique.

Depending of the location and characteristics of the disruption, the transgene construct can be used to generate a transgenic animal in which substantially all expression of the targeted *smoothened* gene is inhibited in at least a portion of the animal's cells. If only regulatory elements are targeted, some low-level expression of the targeted gene may occur (i.e., the targeted allele is "leaky").

The nucleotide sequence(s) comprising the knockout construct(s) can be obtained using methods well known in the art. Such methods include, for example, screening genomic libraries with *smoothened* cDNA probes in order to identify the corresponding genomic *smoothened* gene and regulatory sequences. Alternatively, where the cDNA sequence is to be used as part of the knockout construct, the cDNA may be obtained by screening a cDNA library as set out above.

In another embodiment, the "transgenic non-human animals" of the invention are produced by introducing transgenes into the germline of the non-human animal. Embryonal target cells at various developmental stages can be used to introduce transgenes. Different methods are used depending on the stage of development of the embryonal target cell. The specific line(s) of any animal used to practice this invention are selected for general good health, good embryo yields, good pronuclear visibility in the embryo, and good reproductive fitness. In addition, the haplotype is a significant factor. For example, when transgenic mice are to be produced, strains such as C57BL/6 or FVB lines are often used (Jackson Laboratory, Bar Harbor, ME). Preferred strains are those with H-2^b, H-2^d or H-2^q haplotypes such as C57BL/6 or DBA/1. The line(s) used to practice this invention may themselves be transgenics, and/or may be knockouts (i.e., obtained from animals which have one or more genes partially or completely suppressed).

In one embodiment, the transgene construct is introduced into a single stage embryo. The zygote is the best target for micro-injection. The use of zygotes as a target for gene transfer has a major advantage in that in most cases the injected DNA will be incorporated into the host gene before the first cleavage (Brinster et al. (1985) *PNAS* 82:4438-4442). As a consequence, all cells of the transgenic animal will carry the incorporated transgene. This will in general also be reflected in the efficient transmission of the transgene to offspring of the founder since 50% of the germ cells will harbor the transgene.

Introduction of the transgene nucleotide sequence into the embryo may be accomplished by any means known in the art such as, for example, microinjection, electroporation, or lipofection. Following introduction of the transgene nucleotide sequence into the embryo, the embryo may be incubated *in vitro* for varying amounts of time, or reimplanted into the surrogate host, or both. In vitro incubation to maturity is within the scope of this invention. One common method in to incubate the embryos *in vitro* for about 1-7 days, depending on the species, and then reimplant them into the surrogate host.

Any technique which allows for the addition of the exogenous genetic material into nucleic genetic material can be utilized so long as it is not destructive to the cell, nuclear membrane or other existing cellular or genetic structures. The exogenous genetic material is preferentially inserted into the nucleic genetic material by microinjection. Microinjection of cells and cellular structures is known and is used in the art.

Reimplantation is accomplished using standard methods. Usually, the surrogate host is anesthetized, and the embryos are inserted into the oviduct. The number of embryos implanted into a particular host will vary by species, but will usually be comparable to the number of off spring the species naturally produces.

Transgenic offspring of the surrogate host may be screened for the presence and/or expression of the transgene by any suitable method. Screening is often accomplished by Southern blot or Northern blot analysis, using a probe that is complementary to at least a portion of the transgene. Western blot analysis using an antibody against the protein encoded by the transgene may be employed as an alternative or additional method for screening for the presence of the transgene product. Typically, DNA is prepared from excised tissue and analyzed by Southern analysis or PCR for the transgene. Alternatively, the tissues or cells believed to express the transgene at the highest levels are tested for the presence and expression of the transgene using Southern analysis or PCR, although any tissues or cell types may be used for this analysis.

Retroviral infection can also be used to introduce transgene into a non-human animal. The developing non-human embryo can be cultured *in vitro* to the blastocyst stage. During this time, the blastomeres can be targets for retroviral infection (Jaenich, R. (1976) *PNAS* 73:1260-1264). Efficient infection of the blastomeres is obtained by enzymatic treatment to remove the zona pellucida (*Manipulating the Mouse Embryo*, Hogan eds. (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, 1986). The viral vector system used to introduce the transgene is typically a replication-defective retrovirus carrying the transgene (Jahner et al. (1985) *PNAS* 82:6927-6931; Van der Putten et al. (1985) *PNAS* 82:6148-6152). Transfection is easily and efficiently obtained by culturing the blastomeres on a monolayer of virus-producing cells (Van der Putten, *supra*; Stewart et al. (1987) *EMBO J.* 6:383-388). Alternatively, infection can be performed at a later stage. Virus or virus-producing cells can be injected into the blastocoele (Jahner et al. (1982) *Nature* 298:623-628). Most of the founders will be mosaic for the transgene since incorporation occurs only in a subset of the cells which formed the transgenic non-human animal. Further, the founder may contain various retroviral insertions of the transgene at different positions in the genome which generally will segregate in the offspring. In addition, it is also possible to introduce transgenes into the germ line by intrauterine retroviral infection of the midgestation embryo (Jahner et al. (1982) *supra*).

A third type of target cell for transgene introduction is the embryonal stem cell (ES). ES cells are obtained from pre-implantation embryos cultured *in vitro* and fused with embryos (Evans et al. (1981) *Nature* 292:154-156; Bradley et al. (1984) *Nature* 309:255-258; Gossler et al. (1986) *PNAS* 83: 9065-9069; and Robertson et al. (1986) *Nature* 322:445-448). Transgenes can be efficiently introduced into the ES cells by DNA

transfection or by retrovirus-mediated transduction. Such transformed ES cells can thereafter be combined with blastocysts from a non-human animal. The ES cells thereafter colonize the embryo and contribute to the germ line of the resulting chimeric animal. For review see Jaenisch, R. (1988) *Science* 240:1468-1474.

In one embodiment, gene targeting, which is a method of using homologous recombination to modify an animal's genome, can be used to introduce changes into cultured embryonic stem cells. By targeting the *smoothened* gene in ES cells, these changes can be introduced into the germlines of animals to generate chimeras. The gene targeting procedure is accomplished by introducing into tissue culture cells a DNA targeting construct that includes a segment homologous to a *smoothened* locus, and which also includes an intended sequence modification to the *smoothened* genomic sequence (e.g., insertion, deletion, point mutation). The treated cells are then screened for accurate targeting to identify and isolate those which have been properly targeted.

Gene targeting in embryonic stem cells is in fact a scheme contemplated by the present invention as a means for disrupting a *smoothened* gene function through the use of a targeting transgene construct designed to undergo homologous recombination with *smoothened* genomic sequences. Targeting construct can be arranged so that, upon recombination with an element of a *smoothened* gene, a positive selection marker is inserted into (or replaces) coding sequences of the targeted *smoothened* gene. The inserted sequence functionally disrupts the *smoothened* gene, while also providing a positive selection trait.

Generally, the embryonic stem cells (ES cells) used to produce the knockout animals will be of the same species as the knockout animal to be generated. Thus for example, mouse embryonic stem cells will usually be used for generation of a *smoothened*-knockout mice.

Embryonic stem cells are generated and maintained using methods well known to the skilled artisan such as those described by Doetschman et al. (1985) *J. Embryol. Exp. Morphol.* 87:27-45). Any line of ES cells can be used, however, the line chosen is typically selected for the ability of the cells to integrate into and become part of the germ line of a developing embryo so as to create germ line transmission of the knockout construct. Thus, any ES cell line that is believed to have this capability is suitable for use herein. The cells are cultured and prepared for knockout construct insertion using methods well known to the skilled artisan, such as those set forth by Robertson in: Teratocarcinomas and Embryonic Stem Cells: A Practical Approach, E.J. Robertson, ed. IRL Press, Washington, D.C. [1987]); by Bradley et al. (1986) *Current Topics in Devel. Biol.* 20:357-371); and by Hogan et al. (Manipulating the Mouse Embryo: A Laboratory Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY [1986]) .

Insertion of the knockout construct into the ES cells can be accomplished using a variety of methods well known in the art including for example, electroporation, microinjection, and calcium phosphate treatment. A preferred method of insertion is electroporation .

Each knockout construct to be inserted into the cell must first be in the linear form. Therefore, if the knockout construct has been inserted into a vector, linearization is accomplished by digesting the DNA with a suitable restriction endonuclease selected to cut only within the vector sequence and not within the knockout construct sequence.

For insertion, the knockout construct is added to the ES cells under appropriate conditions for the insertion method chosen, as is known to the skilled artisan. Where more than one construct is to be introduced into the ES cell, each knockout construct can be introduced simultaneously or one at a time.

If the ES cells are to be electroporated, the ES cells and knockout construct DNA are exposed to an electric pulse using an electroporation machine and following the manufacturer's guidelines for use. After electroporation, the ES cells are typically allowed to recover under suitable incubation conditions. The cells are then screened for the presence of the knockout construct .

Screening can be accomplished using a variety of methods. Where the marker gene is an antibiotic resistance gene, the ES cells may be cultured in the presence of an otherwise lethal concentration of antibiotic. Those ES cells that survive have presumably integrated the knockout construct. If the marker gene is other than an antibiotic resistance gene, a Southern blot of the ES cell genomic DNA can be probed with a sequence of DNA designed to hybridize only to the marker sequence Alternatively, PCR can be used. Finally, if the marker gene is a gene that encodes an enzyme whose activity can be detected (e.g., β-galactosidase), the enzyme substrate can be added to the cells under suitable conditions, and the enzymatic activity can be analyzed. One skilled in the art will be familiar with other useful markers and the means for detecting their presence in a given cell. All such markers are contemplated as being included within the scope of the teaching of this invention.

The knockout construct may integrate into several locations in the ES cell genome, and may integrate into a different location in each ES cell's genome due to the occurrence of random insertion events. The desired location of insertion is in a complementary position to the DNA sequence to be knocked out, e.g., the *smoothened* coding sequence, transcriptional regulatory sequence, etc. Typically, less than about 1-5 percent of the ES cells that take up the knockout construct will actually integrate the knockout construct in the desired location. To identify those ES cells with proper integration of the knockout construct, total DNA can be extracted from the ES cells using standard methods. The DNA can then be probed on a

Southern blot with a probe or probes designed to hybridize in a specific pattern to genomic DNA digested with particular restriction enzyme(s). Alternatively, or additionally, the genomic DNA can be amplified by PCR with probes specifically designed to amplify DNA fragments of a particular size and sequence (i.e., only those cells containing the knockout construct in the proper position will generate DNA fragments of the proper size).

After suitable ES cells containing the knockout construct in the proper location have been identified, the cells can be inserted into an embryo. Insertion may be accomplished in a variety of ways known to the skilled artisan, however a preferred method is by microinjection. For microinjection, about 10-30 cells are collected into a micropipet and injected into embryos that are at the proper stage of development to permit integration of the foreign ES cell containing the knockout construct into the developing embryo. For instance, the transformed ES cells can be microinjected into blastocytes.

After the ES cell has been introduced into the embryo, the embryo may be implanted into the uterus of a pseudopregnant foster mother for gestation. While any foster mother may be used, the foster mother is typically selected for her ability to breed and reproduce well, and for her ability to care for the young. Such foster mothers are typically prepared by mating with vasectomized males of the same species. The stage of the pseudopregnant foster mother is important for successful implantation, and it is species dependent.

Offspring that are born to the foster mother may be screened initially for *smoothened* disruptants, DNA from tissue of the offspring may be screened for the presence of the knockout construct using Southern blots and/or PCR as described above. Offspring that appear to be mosaics may then be crossed to each other, if they are believed to carry the knockout construct in their germ line, in order to generate homozygous knockout animals. Homozygotes may be identified by Southern blotting of equivalent amounts of genomic DNA from animals that are the product of this cross, as well as animals that are known heterozygotes and wild type animals.

Other means of identifying and characterizing the knockout offspring are available. For example, Northern blots can be used to probe the mRNA for the presence or absence of transcripts of either the *smoothened* gene, the marker gene, or both. In addition, Western blots can be used to assess the (loss of) level of expression of the *smoothened* gene knocked out in various tissues of the offspring by probing the Western blot with an antibody against the *smoothened* protein, or an antibody against the marker gene product, where this gene is expressed. Finally, *in situ* analysis (such as fixing the cells and labeling with antibody) and/or FACS (fluorescence activated cell sorting) analysis of various cells from the offspring can be conducted using suitable antibodies or *smoothened* ligands, e.g., *hedgehog* proteins, to look for the presence or absence of the knockout construct gene product.

Animals containing more than one knockout construct and/or more than one transgene expression construct are prepared in any of several ways. The preferred manner of preparation is to generate a series of animals, each containing a desired transgenic phenotypes. Such animals are bred together through a series of crosses. backcrosses and selections, to ultimately generate a single animal containing all desired knockout constructs and/or expression constructs, where the animal is otherwise congenic (genetically identical) to the wild type except for the presence of the knockout construct(s) and/or transgene(s). Thus, a transgenic avian species can be generated by breeding a first transgenic bird in which the wild-type *smoothened* gene is disrupted with a second transgenic bird which has been engineered to express a mutant *smoothened* which retains most other biological functions of the receptor.

The transformed animals, their progeny, and cell lines of the present invention provide several important uses that will be readily apparent to one of ordinary skill in the art.

To illustrate, the transgenic animals and cell lines are particularly useful in screening compounds that have potential as prophylactic or therapeutic treatments of diseases such as may involve aberrant expression, or loss, of a *smoothened* gene, or aberrant or unwanted activation of receptor signaling. Screening for a useful drug would involve administering the candidate drug over a range of doses to the transgenic animal, and assaying at various time points for the effect(s) of the drug on the disease or disorder being evaluated. Alternatively, or additionally, the drug could be administered prior to or simultaneously with exposure to induction of the disease, if applicable.

In one embodiment, candidate compounds are screened by being administered to the transgenic animal, over a range of doses, and evaluating the animal's physiological response to the compound(s) over time. Administration may be oral, or by suitable injection, depending on the chemical nature of the compound being evaluated. In some cases, it may be appropriate to administer the compound in conjunction with co-factors that would enhance the efficacy of the compound.

In screening cell lines derived from the subject transgenic animals for compounds useful in treating various disorders, the test compound is added to the cell culture medium at the appropriate time, and the cellular response to the compound is evaluated over time using the appropriate biochemical and/or histological assays. In some cases, it may be appropriate to apply the compound of interest to the culture medium in conjunction with co-factors that would enhance the efficacy of the compound.

Exemplification

The invention now being generally described, it will be more readily understood by reference to the following examples which are included merely for purposes of illustration of certain aspects and embodiments of the present invention, and are not intended to limit the invention.

In the *Drosophila* embryo, spatially restricted expression of the *wingless* (*wg*) signaling protein, which is essential for the normal patterning of each segment (reviewed in Klingensmith, J. & Nusse, R. (1994) *Dev. Biol.* 166:3966-414), is controlled by the localized activity of *hedgehog* (*hh*) in neighboring cells (reviewed in Ingham, P.W. (1995) *supra*). Expression of *hh* is in turn maintained by *wg* activity, and these mutual regulatory interactions stabilize the expression domains of the two genes and hence the parasegment boundaries defined by their interfaces (reviewed in Ingham, P.W. & Martinas-Arias, A. (1992) *Cell* 68:221-235). In the absence of either gene activity, these boundaries disappear and all patterning and polarity of the segments is lost (Fig. 1 e, f). Embryos homozygous for mutations of the *smoothened* (*smo*) locus (formerly named *smooth*) (Nüsslein-Volhard, C. et al. (1984) *Roux's Arch. Dev. Biol.* 193:267-282) exhibit phenotypes similar to those of weak alleles of *hh* or *wg*, affecting the patterning of each segment in a highly variable manner (Fig. 1a-c). In all cases, this phenotype is stronger at 18 ° C than at 25 ° C, suggesting that the variability may be due to the hypomorphism of each mutant *smo* allele. However, such variability might also be accounted for if maternally derived *smo* partially compensates for the loss of zygotic gene activity. To investigate this possibility, we generated mosaic females lacking wild-type *smo* alleles in their germ line (smo_{glc}): *smo* germline clones) and analyzed the phenotype of homozygous *smo* embryos from such smo_{glc} females (hereafter designated *smo-* embryos). These *smo-* embryos display an invariant phenotype almost indistinguishable from that of *hh* null alleles (Fig. 1d-f): thus significant amounts of *smo* product are contributed to the egg during oogenesis, although this contribution is neither sufficient (as *smo* homozygous from heterozygous females die) nor necessary (as *smo* heterozygotes derived from smo_{glc} females survive) fir normal development.

The strong similarity between the *hh* null and the *smo* phenotypes is suggestive of a role for *smo* in *hh* signaling, and consistent with this, *smo-* embryos lack *wg* transcription at their parasegment boundaries (Fig 2a, d). As *wg* transcription also disappears in *wg* mutant embryos (Bejsovec, A. & Martinez-Arias, A. (1991) *Development* 113:471-485; Ingham, P.W. & Hidalgo, A. (1993) *Development* 117:283-291; van den Huevel, M. et al. (1993) *EMBO J* 12:5293-5302), however, this finding does not in itself exclude a role for *smo* in *wg* signaling. To address this issue directly, we examined the ability of each signal too function in the absence of *smo* activity by expressing either gene ectopically under the control of heterologous regulatory elements in *smo-* embryos. Expression of *wg* in alternate parasegments of a *smo-* embryo results in the partial restoration of posterior naked cuticle in

alternate segments (Fig. 1g); by contrast, expression of *hh* under identical conditions has no effect on the cuticular phenotype of a *smo*- embryo (Fig. 1h). These findings suggest that *smo* is required for the activity of *hh* but not of *wg*. To confirm this inference, the expression of *engrailed* (*en*) and *wg*, targets of *wg* and *hh* activity respectively, were analyzed in each type of embryo. Whereas in a wild-type background, misexpression of *hh* induces ectopic *wg* expression in *smo* embryos, such *hh* misexpression has no effect on *wg* transcription (Fig. 2b, c) as would be expected if *smo* is required for *hh* signaling. In contrast, ectopic *wg* expression in *smo*- embryos restores the expression of *en* (Fig. 2g-i), confirming that *wg* signaling can occur in the absence of *smo* activity. Interestingly, however, transcription of the endogenous *wg* gene is not itself restored (Fig. 2e, f), implying that, contrary to previous assertions (Bejsovec, A. & Martinas-Arias, A. (1991) *supra*; Ingham, P.W. & Hidalgo, A. (1993) *supra*; van den Huevel, M. et al. (1993) *supra*; Hooper, J.E. (1994) *Nature* 372:461-464), *wg* activity alone is not sufficient to maintain its own expression.

In the developing imaginal discs, expression of *hh* is restricted to cells of the posterior lineage compartment and, as in the embryo, its activity is required by neighboring anterior compartment cells for the transcription of other signal-encoding genes (reviewed in Blair, S.S. (1995) *BioEssays* 17:299-309). In the case of the wing disc, the principal target of *hh* activity is the decapentaplegic (*dpp*) gene, expression of which is restricted to a thin stripe of cells running along the anterior side of the compartment boundary (Raftery, L.A. et al. (1991) *Development* 113:27-33) (Fig 3a). To determine whether *smo* is required for this *hh*-dependent expression of *dpp*, we used flippase-induced somatic recombination (Xu, T. & Rubin, G.M. (1993) *Development* 117:1223-1237) to generate mosaic imaginal discs in which small clones of cells lack wild-type *smo* activity. When such clones are located anterior to the stripe of *dpp*-expressing cells or anywhere in the posterior compartment (Fig 3a?), the expression of *dpp* along the border is unaffected.

Expression of *dpp* is lost, however, from clones of cells lacking *smo* activity when such clones are located within the normal *dpp* domain (Fig. 3b-e). Thus *smo* is required in a cell-autonomous manner for *dpp* expression. Transcription of *dpp* can be activated in the wing disc independently of *hh* by removing the activity of cAMP-dependent protein kinase (PKA), suggesting that PKA acts downstream of *hh* to antagonize its activity (Kalderon, D. (1995) *Curr. Biol.* 5:580-582). To determine whether such *hh* independent expression of *dpp* requires the activity of *smo*, we generated clones that simultaneously lack *smo* and the PKA catalytic subunit. Such clones express *dpp* at any position within the anterior compartment of the disc (Fig. 3f, g), indicating that *smo* is not absolutely required for *dpp* transcription; rather, it acts upstream of PKA to mediate activation of *dpp* by *hh*.

We mapped the *smo* locus more precisely by recombination analysis and by using small chromosomal deficiencies, localized it to bands 21B7-8 on the left arm of chromosome II. Genomic P1 clones (Smoller, D.A. et al. (1991) *Chromasoma* 100:487-494) covering this region were obtained and used to screen genomic DNA from newly isolated -ray-induced alleles (Fig. 1) of *smo* by Southern blot hybridization. We identified a fragment contained within one of the P1 clones that consistently detects a novel restriction fragment in digests from the -ray-induced allele smo^{D16} (Fig. 4a). Northern blot analysis of RNA from adult females revealed three maternally expressed transcripts recognized by this fragment (Fig. 4c). The largest of these (4.1kilobases) shows a size alternation in RNA from smo^{D16} heterozygous females (Fig. 4b), suggesting that it may represent the *smo* transcript. To investigate this possibility, a genomic fragment including only this transcription unit (Fig. 4c) as transformed into flies and tested for its ability to rescue the *smo* phenotype. Animals homozygous for *smo* alleles and carrying one copy of this transgene survive to adulthood and are phenotypically wild type, confirming that this fragment includes the *smo* transcription unit and all sequences necessary for its expression (data not shown).

The complete nucleotide sequence of the *smo* transcription unit was determined by sequencing genomic and partial complementary DNA clones (Fig. 4d) to reveal the intron-exon structure shown in Fig. 4c. Conceptual translation of the sequence identified a single long open reading frame of 1,024 amino acids (Fig. 4d) preceded by a 5' untranslated leader of at least 296 nucleotides. Downstream of the translation termination codon is an untranslated region of 654 nucleotides. Hydropathy analysis of the amino acid sequence indicates a putative signal sequence followed by a further seven hydrophobic domains, each of which is long enough to span the membrane once. These features are typical of members of the large family of G-protein-coupled receptors (GPCR); searches of the databases revealed no significant homology to any well characterized GPCR, but did reveal limited homology to members of the *Frizzled* (*Fz*) family of serpentine proteins (Wang, Y. et al. (1996) *J. Biol. Chem.* 271:4468-4476) (Fig. 4d). In contrast to the latter, the predicted *smo* protein contains a long carboxy-terminal extension, which includes consensus target sites for PLA and G-protein-coupled-receptor kinases (Fig. 4d), both of which are involved in GPCR desensitization after agonist-induced activation (Dohlman, H.G. et al. (1991) *Annu. Rev. Biochem.* 60:653-688). On the basis of these features *smo* may act as a receptor for the *hh* protein. The involvement of both PKA and a protein with characteristics of GPCR in transduction of the Hh signal is intriguing, given the well established roles of several GPCRs in modulating adenylyl cyclase activity and hence intracellular cyclic AMP levels (Dohlman, H.G. et al. (1991) *supra*). We note, however, that PKA may not be directly regulated by *hh* activity, but rather may act in a parallel pathway to antagonize the targets of *hh* signaling (Kalderon, D. (1995) *supra*). Moreover, recent studies have emphasized the

diversity of pathways by which the activity of GPCR agonists can be transduced (Daub, H. et al. (1996) *Nature* 379:557-560; Grand, R.J.A. et al. (1996) *Biochem. J.* 313:353-368). Earlier analysis has implicated another multiple membrane-spanning protein, encoded by the segment-polarity gene *patched* (*ptc*) (Nakano, Y. et al. (1989) *Nature* 341:508-513; Hooper, J. & Scott, M.P. (1989) *Cell* 59:751-765), in the reception of the Hh signal (Ingham, P.W. et al. (1991) *supra*). Although our results cast doubt on models of Ptc as the Hh receptor, we emphasize that the functional relationship between Hh, Ptc and Smo remains unclear. As Ptc has a predicted topology reminiscent of proteins involved in transport of ions and small molecule (Nakano, Y. et al. (1989) *supra;* Hooper, J. & Scott, M.P. (1989) *supra*), its activity may be regulated by Smo in a manner analogous to that of G-protein-gated ion channels (Clapham, D.E. & Neer, E.J. (1993) *Nature* 365:508-513). Epistasis analysis has, however, placed *ptc* upstream of *smo* (Hooper, J.E. (1994) *supra*), suggesting that there is an unprecedented relationship between these two unusual proteins, the nature of which awaits further analysis.

Utilizing the Drosophila *smoothened* coding sequence, we have also identified vertebrate homologs of the *smoothened* gene. A nearly complete chicken sequence was cloned and its sequence is given in SEQ ID No. 4 (nucleotide) and 8 (protein).

All of the above-cited references and publications are hereby incorporated by reference.

We Claim:

1. An isolated and/or recombinant *smoothened* polypeptide comprising a *smoothened* amino acid sequence identical or homologous to an amino acid sequence represented in any of SEQ ID Nos. 5, 6, 7 or 8.

2. An isolated and/or recombinant *smoothened* polypeptide comprising a *smoothened* amino acid sequence at least 63 percent similar to the amino acid sequence represented in SEQ ID No. 5, 6, 7 or 8, or a portion thereof, and, which polypeptide interacts with a *patched* protein.

3. An isolated and/or recombinant *smoothened* polypeptide comprising an amino acid sequence encoded by a nucleic acid which hybridizes under stringent conditions to a metazoan *smoothened* gene.

4. An isolated and/or recombinant *smoothened* polypeptide comprising an amino acid sequence cross-reactive with an antibody specific for the *smoothened* protein designated in SEQ ID No. 5, 6, 7 or 8, which *smoothened* polypeptide is specifically interacts with a *patched* protein.

5. The *smoothened* polypeptide of any of claims 1, 2, 3 or 4, which polypeptide modulates at least one of proliferation, differentiation or survival of a cell which expresses the *smoothened* polypeptide.

6. The *smoothened* polypeptide of claim 5, wherein the cell is a neuronal cell.

7. The *smoothened* polypeptide of claim 5, wherein the cell is a osteogenic or chrondocytic cell.

8. The *smoothened* polypeptide of claim 3, wherein the cell is a testicular cell.

9. The *smoothened* polypeptide of any of claims 1, 2, 3 or 4, which polypeptide comprises an amino acid sequence at least 75% homologous with the amino acid sequence designated by SEQ ID No: 5, 6, 7 or 8.

10. The *smoothened* polypeptide of claim 9, which polypeptide comprises an amino acid sequence at least 85% homologous with the amino acid sequence designated by SEQ ID No: 5, 6, 7 or 8.

11. The *smoothened* polypeptide of claim 9, which polypeptide comprises an amino acid sequence at least 95% homologous with the amino acid sequence designated by SEQ ID No: 5, 6, 7 or 8.

12. The *smoothened* polypeptide of claim 9, which polypeptide comprises an amino acid sequence identical with the amino acid sequence designated by SEQ ID No: 5, 6, 7 or 8.

12. The *smoothened* polypeptide of claim 3, wherein the *smoothened* gene includes an open reading frame designated by SEQ ID No: 1, 2, 3, 4 or 9.

13. The *smoothened* polypeptide of any of claims 1, 2, 3 or 4, which polypeptide comprises a seven transmembrane domains, at least one N-linked glycosylation in an extracellular domain, at least one phosphorylation sites for a cAMP-dependent kinases in an intracellular domain.

14. The *smoothened* polypeptide of any of claims 1, 2, 3 or 4, which polypeptide is encoded by a *smoothened* gene of mammalian origin.

15. An immunogen comprising the *smoothened* polypeptide of claim 1, in an immunogenic preparation, the immunogen being capable of eliciting an immune response specific for the *smoothened* polypeptide.

16. An antibody preparation specifically reactive with an epitope of the *smoothened* polypeptide of claim 1.

17. An isolated nucleic acid comprising a coding sequence encoding a recombinant polypeptide comprising a *smoothened* polypeptide sequence identical or homologous to an amino acid sequence represented in SEQ ID No. 5, 6, 7 or 8.

18. An isolated nucleic acid encoding a recombinant polypeptide comprising a *smoothened* coding sequence which hybridizes to a metazoan *smoothened* gene.

19. The nucleic acid of any of claims 17 or 18, which coding sequence hybridizes under stringent conditions to a nucleic acid probe having a sequence represented by at least 12 consecutive nucleotides of SEQ ID No. 1, 2, 3 or 4.

20. The nucleic acid of any of claims 17 or 18, further comprising a transcriptional regulatory sequence operably linked to the coding sequence so as to render the nucleic acid suitable for use as an expression vector.

21. An expression vector, capable of replicating in at least one of a prokaryotic cell and eukaryotic cell, comprising the nucleic acid of claim 20.

22. A host cell transfected with the expression vector of claim 21 and expressing the recombinant polypeptide.

23. A method of producing a recombinant *smoothened* polypeptide comprising culturing the cell of claim 22 in a cell culture medium to cause expression of a *smoothened* polypeptide encoded by the expression vector, and isolating the *smoothened* polypeptide from the cell culture.

24. A transgenic animal having cells which harbor a transgene comprising the nucleic acid of claim 17.

25. A transgenic animal in which *smoothened* stimulated signal transduction pathways are inhibited in one or more tissue of the animal by one of either expression of an antagonistic *smoothened* polypeptide or disruption of a *smoothened* gene.

26. A recombinant gene comprising a *smoothened* encoding nucleotide sequence identical or homologous with SEQ ID No. 1, 2, 3 or 4, or a fragment thereof, the nucleotide sequence operably linked to a transcriptional regulatory sequence in an open reading frame and translatable to a polypeptide capable of specifically regulating *hedgehog* signal transduction.

27. The recombinant gene of claim 26, wherein the *smoothened* encoding nucleotide sequence is derived from a cDNA clone.

28. The recombinant gene of claim 26, wherein the *smoothened* encoding nucleotide sequence is derived from a genomic clone and includes intronic nucleotide sequences disrupting the open reading frame.

29. A nucleic acid comprising a substantially purified oligonucleotide, the oligonucleotide containing a region of nucleotide sequence which hybridizes under stringent

conditions to at least 10 consecutive nucleotides of sense or antisense sequence of SEQ ID No. 1, 2, 3, 4 or 9, or naturally occurring mutants thereof.

30. The nucleic acid of claim 29, which nucleic acid further comprises a label group attached thereto and able to be detected.

31. A test kit for detecting cells which contain a *smoothened* mRNA transcript, comprising a nucleic acid of claim 29 for measuring, in a sample of cells, a level of nucleic acid encoding a *smoothened* protein.

32. A test kit for detecting cells or tissue containing a *smoothened* protein, comprising an antibody specific for a *smoothened protein* for measuring, in a sample of cells, a level of the *smoothened* protein.

33. A method for modulating, in an animal, cell growth, differentiation or survival, comprising administering a therapeutically effective amount of a *smoothened* polypeptide which modulates *hedgehog*-dependent signal transduction.

34. The method of claim 33, comprising administering a nucleic acid construct encoding a *smoothened* polypeptide under conditions wherein the construct is incorporated and recombinantly expressed by the cells to be modulated or cells located proximate thereto.

35. The method of claim 33, comprising administering a peptidomimetic of a *smoothened* protein, which peptidomimetic binds to and modulates intracellular signal transduction pathways mediated by *patched*.

36. A recombinant transfection system, comprising
    (i) a gene construct encoding a *smoothened* polypeptide and operably linked to a transcriptional regulatory sequence for causing expression of the *smoothened* polypeptide in eukaryotic cells, and
    (ii) a gene delivery composition for delivering the gene construct to a cell and causing the cell to be transfected with the gene construct.

37. The recombinant transfection system of claim 36, wherein the gene delivery composition is selected from a group consisting of a recombinant viral particle, a liposome, and a poly-cationic nucleic acid binding agent,

38. A method of determining if a subject is at risk for a disorder characterized by unwanted cell proliferation, differentiation or death, comprising detecting, in a tissue of the subject, the presence or absence of a genetic lesion characterized by at least one of (i) a mutation of a gene encoding a *smoothened* protein; and (ii) the mis-expression of the gene.

39. The method of claim 38, wherein detecting the genetic lesion comprises ascertaining the existence of at least one of
    i. a deletion of one or more nucleotides from the gene,
    ii. an addition of one or more nucleotides to the gene,
    iii. an substitution of one or more nucleotides of the gene,
    iv. a gross chromosomal rearrangement of the gene,
    v. aberrant methylation of the gene,
    vi. a gross alteration in the level of a messenger RNA transcript of the gene,
    vii. the presence of a non-wild type splicing pattern of a messenger RNA transcript of the gene, and
    viii. a non-wild type level of the protein.

40. The method of claim 39, wherein detecting the genetic lesion comprises
    i. providing a nucleic acid comprising an oligonucleotide containing a region of nucleotide sequence which hybridizes to a sense or antisense sequence of SEQ ID No. 1, 2, 3, 4 or 9, or naturally occurring mutants thereof or 5' or 3' flanking sequences naturally associated with the gene;
    ii. exposing the nucleic acid to nucleic acid of the tissue; and
    iii. detecting, by hybridization of the nucleic acid to the nucleic acid, the presence or absence of the genetic lesion.

41. The method of claim 39, wherein detection of the genetic lesion comprises detecting the presence or absence of a *smoothened* protein in cells of a tissue sample and/or as soluble proteins in bodily fluid.

42. An assay for screening test compounds that modulate the bioactivity of a *smoothened* receptor comprising:
    i. combining a test compound, a *smoothened* polypeptide, and a target compound selected from the group consisting of a *smoothened* ligand, a signal transduction protein which binds to the *smoothened* polypeptide, or *patched* protein; and

ii. detecting the interaction of the target compound and the *smoothened* polypeptide,

wherein a change in the interaction of the target compound and the *smoothened* polypeptide in the presence of the test compound is indicative of a potential ability to modulate the bioactivity of the *smoothened* receptor.

43. A peptidomimetic of a portion of a *smoothened* protein which specifically binds to one of a *smoothened* ligand, a signal transduction protein or *patched*, and modulates *hedgehog*-mediated signal transduction of the *smoothened* protein.

44. A method for causing expression of a *smoothened* protein, comprising transfecting a cell with a gene activation construct which recombines with a genomic *smoothened* gene of the patient to provide a heterologous transcriptional regulatory sequence operatively linked to a coding sequence of the *smoothened* gene.

45. A gene activation construct comprising

(i) a targeting sequence including a nucleic acid sequence which is substantially identical to or substantially complementary to a genomic *smoothened* gene sequence, which targeting sequence is sufficient to cause homologous recombination between a genomic *smoothened* gene and the targeting transgene construct, and

(ii) a transcriptional regulatory sequence,

wherein homologous recombination of the targeting sequence with a genomic *smoothened* gene disposes the transcriptional regulatory sequence in operative control of expression of the genomic *smoothened* gene.

FIG. I

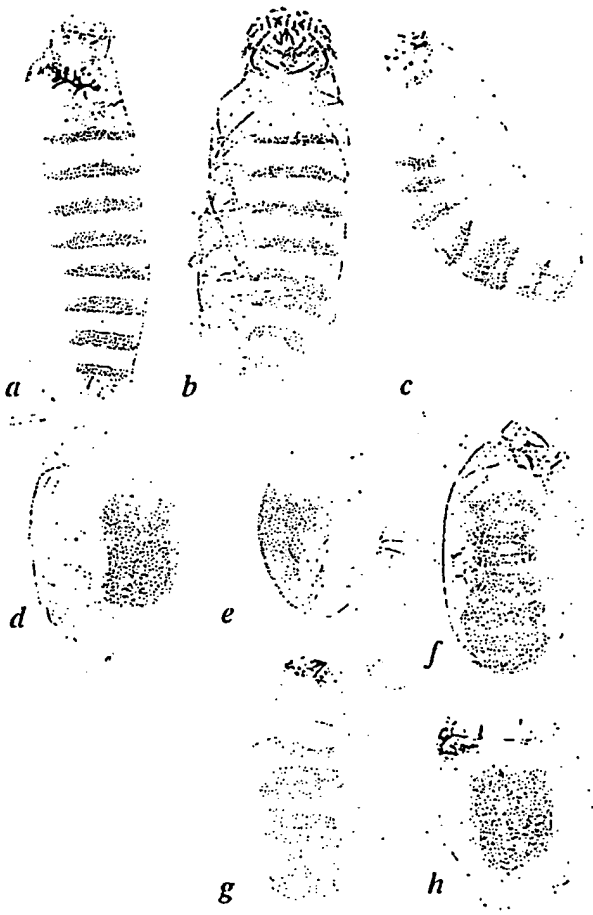


FIG. II

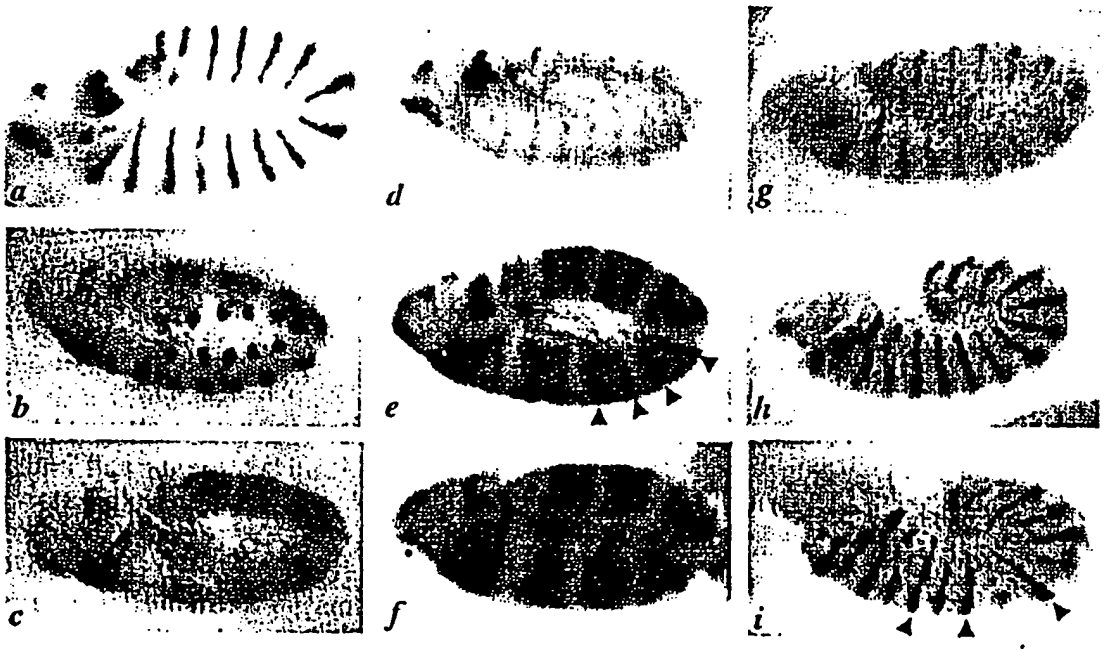


FIG. III

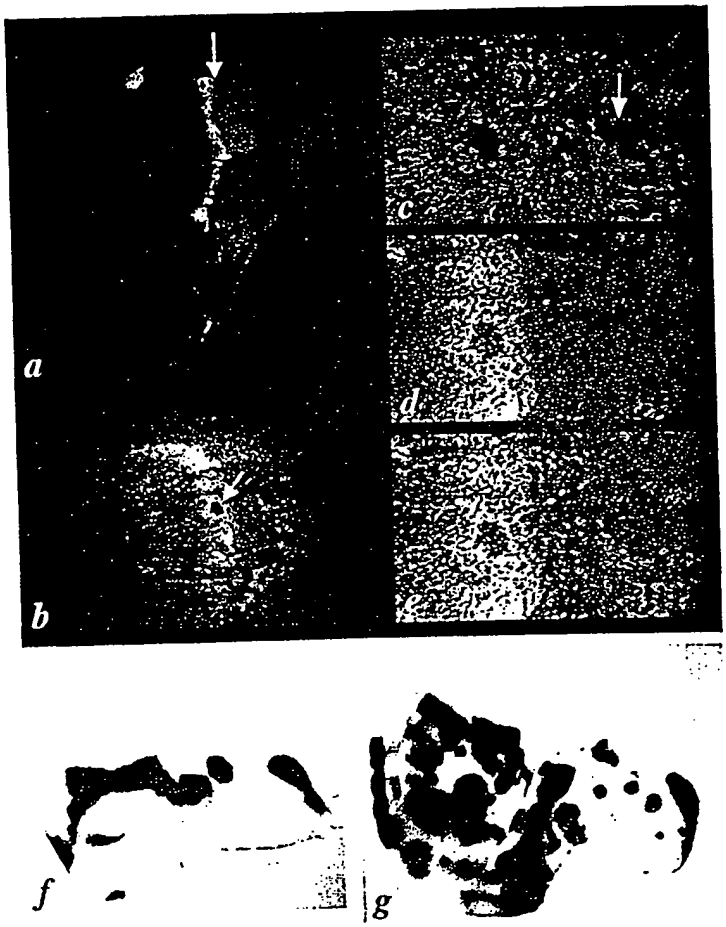
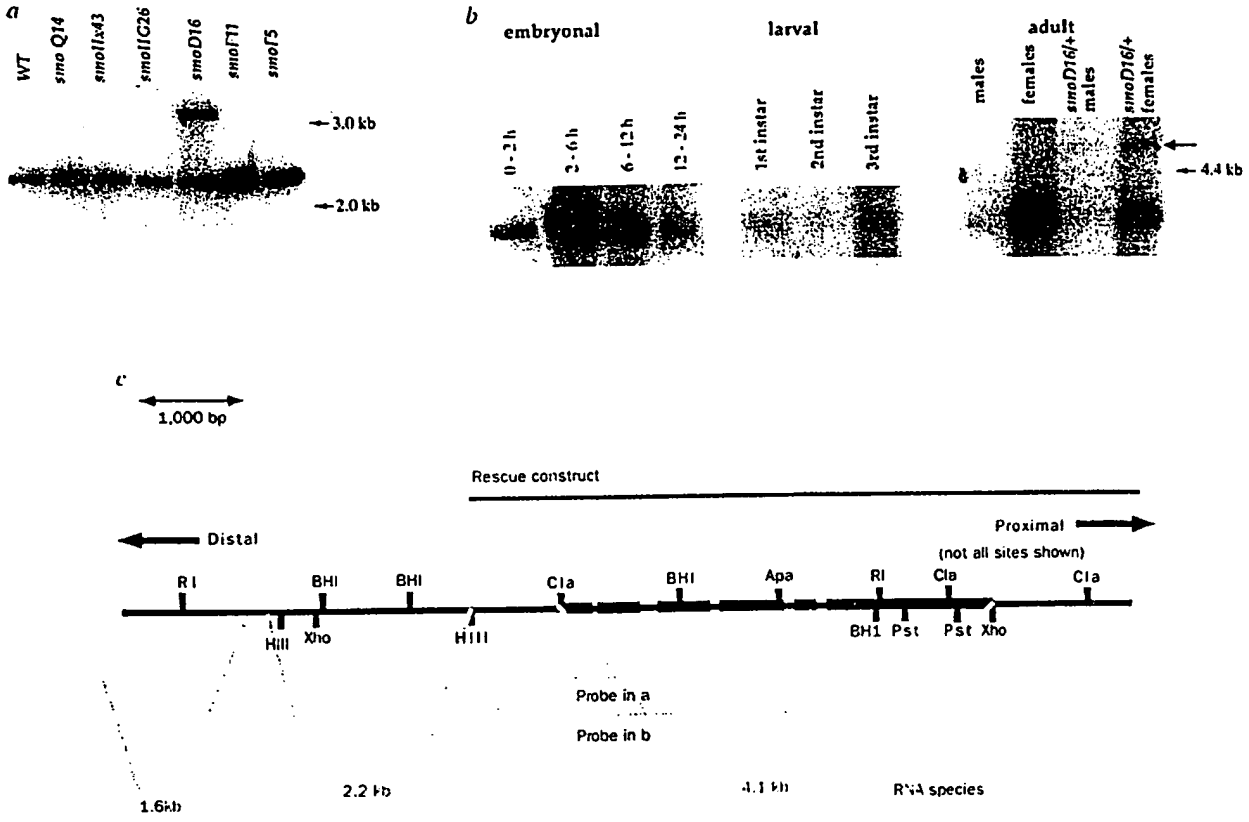


FIG. IV



d

```
MNFLEVAILC LWVVADASAG SAFFGSTTPA SAQQSDVELE PINGTLNYRL   50
YAKKGRDDKP WFDGLDSPHI QCVRRARCYP TSINAINTCFG SKLPYELSSL  100
DLTDFHTEKE LNDKLNDYYA LKHVPKCWAA IQPFLCAVFK PKCEKINGED  150
HVTLPSYEMC RITMEPCRIL YNTTFFPKFL RCNMTLFPTK CTNGARGMKF  200
NGTGQCLSPL VPTDTSASYY PGIEVCGVRC KDPLYTDDEH RQIHKLIGWA  250
GSICLLSNLF VVSTFFIDWK NANKYPAVIV FYINLCFLIA CVGWLLQFTS  300
GSREDIVCRK DGTLRHSEPT AGENLSCIVI FVLVYYFLTA GMVWFVFLTY  350
AMHNRANGHV QDRIDKKGSY FHLVANSLPL VLTITTMAFS EVDGNSIVGI  400
CFVGYINHGM RAGLLLGPLC GVILSGGYFI TRGMVMLFGL KHFANDIKST  450
SASNKIHLII MRMGVCALLT LVFILVAIAC HVTEFRHADE WAQSFRQFII  500
CRISSVFEEK SSCRIEHRPS VGVLQLHLLC LFSSGIVMST WCWTPSSIET  550
WKRYIRKKCG KEVVEEVKNP KHKVIAQTWA KRKDFEDKGR LSITLYNTHT  600
DPYGLNFDVN DLNSSETNDI SSTWAAYLPQ CVKRRMALTG AATGNSSIHG  650
PPKNSLDSEI SVSVRHVSVE SRRNSVDSQV SVKIAEMKTK VASRSRGKHG  700
GDSENRRTQR RRDYIAAATG KSSRRRRSST SVESQVIALK KTTYPNASHK  750
VGYFAHHSSK KQHNYTSSMK RRTANAGLDP SILNEFLQKN GDFIFPFLQN  800
QDNSSSNEED NSRASQKIQD LNYVVKQQEI SEDDHDGIKI EELPNSKQVA  850
LENFLKNIKK SNESNSNRHS RNEARSQSKK SQKRHLKNPA ADLDFRKDCV  900
KYRSNDSLSC SSEELDVALD VGSLLNSSFS GISMGKPHSR NSKTSCDVGI  950
QANIPFELVPS YGEDELQQAM RLLNAASRQR TEAANEDFGG TELQGLLGHS 1000
HRHQREPTFM SESDKLKMLL LPEQ                               1024
```

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT:
(A) NAME: ONTOGENY, INC.
(B) STREET: 45 Moulton Street
(C) CITY: Cambridge
(D) STATE: Massachusetts
(E) COUNTRY: United States of America
(F) POSTAL CODE (ZIP): 02138

(ii) TITLE OF INVENTION: Vertebrate Smoothened Gene, Gene Products, and Uses Related Thereto

(iii) NUMBER OF SEQUENCES: 9

(iv) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: AscII(text)

(vi) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER: US 08/897,798
(B) FILING DATE: July 21, 1997

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 4039 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: both
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 271..3378

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

GGGTATATTT AAACTGGCCC TGAATGGTCG CACATTTGTT GCTTCAGCTT CACGTTGAGT 60

CGCGTTTATT TTTGTTTTCC TCCTTGGTTC TTGTTTTTTT TTTCTCTTTT TTAAAAACAC 120

TGCAATCCTT CAAAGGGAAC AATTAAACCA GGATATCCCG AGTCCAACGA AAGCGCCTTC 180

ACGATGCACT GACGAACCGC TGAGCAAGAA ACGGGCATTC AGCCCATTAA AATTCACACA 240

AGGCAGCTAT TTAATGATTT TTATAAGCGG ATGCAGTACT TAAACTTTCC GCGCATGCCA 300

AACATTATGA TGTTCCTGGA GGTTGCGATC TTATGCCTGT GGGTGGTCGC AGACGCATCG 360

GCCAGTTCGG CCAAGTTCGG CAGCACAACG CCCGCAAGTG CGCAGCAGTC GGATGTGGAA 420

CTGGAGCCCA TCAATGGGAC TCTCAATTAC CGACTGTACG CCAAGAAAGG CAGGGACGAC 480

AAACCCTGGT TTGATGGCCT AGACAGCAGG CACATCCAGT GTGTCCGACG TGCCCGTTGC 540

TACCCCACCT CGAACGCAAC CAACACCTGT TTCGGCTCAA AATTGCCCTA TGAGCTGAGC 600

AGCCTAGATC TCACCGACTT CCACACCGAA AAGGAGCTGA ACGATAAGCT GAACGACTAC 660

TATGCCCTGA AGCACGTGCC CAAATGTTGG GCAGCTATAC AGCCCTTTTT GTGCGCCGTC 720

TTTAAGCCGA AGTGTGAAAA AATCAACGGC GAGGACATGG TCTACCTGCC ATCTTACGAG 780

ATGTGCCGAA TTACCATGGA ACCCTGTCGC ATTTTGTACA ACACGACGTT TTTCCCAAAA 840

TTCCTTCGCT GCAACGAAAC ACTCTTTCCG ACGAAATGCA CAAACGGAGC ACGAGGAATG 900

AAATTCAACG GAACTGGCCA GTGTTTAAGT CCTTTGGTTC CGACAGATAC GTCAGCCAGC 960

TATTATCCTG GCATCGAGGG CTGCGGCGTG CGATGCAAGG ATCCACTCTA TACCGATGAT 1020

GAGCATCGGC AGATCCACAA ACTGATCGGA TGGGCTGGCA GCATATGTCT TCTGTCTAAC 1080

CTTTTCGTGG TGTCCACCTT CTTCATCGAC TGGAAGAATG CCAACAAGTA TCCGGCAGTA 1140

ATTGTGTTCT ACATAAATCT TTGCTTTCTA ATTGCTTGCG TCGGCTGGTT GCTTCAGTTT 1200

ACTTCTGGCT CGCGAGAGGA CATAGTATGT CGTAAGGATG GAACACTTCG CCACTCAGAG 1260

CCTACAGCCG GTGAAAATCT TTCTTGCATA GTGATCTTTG TGCTGGTCTA TTATTTTCTC 1320

ACCGCTGGAA TGGTTTGGTT TGTGTTCCTC ACCTACGCCT GGCATTGGAG GGCCATGGGC 1380

CACGTCCAAG ATCGGATAGA TAAGAAAGGT TCCTACTTTC ACCTCGTGGC GTGGTCACTA 1440

CCCCTTGTGC TTACCATTAC CACGATGGCT TTCAGTGAGG TGGATGGAAA TAGTATTGTG 1500

GGCATCTGCT TCGTAGGCTA TATCAATCAT TCTATGAGGG CAGGACTACT TCTTGGTCCG 1560

CTCTGCGGGG TCATCCTCAT TGGTGGATAC TTCATCACCC GCGGCATGGT GATGCTTTTT 1620

GGACTGAAAC ACTTCGCTAA TGACATTAAA TCAACTTCGG CGAGCAACAA AATCCATTTG 1680

ATCATCATGC GCATGGGAGT CTGTGCTCTG CTCACTTTAG TTTTCATACT AGTGGCCATT 1740

GCGTGCCACG TTACGGAGTT TAGGCATGCA GACGAATGGG CCCAGAGCTT CAGACAGTTT 1800

ATAATCTGCA AAATTTCTTC AGTTTTTGAA GAAAAGAGTT CCTGTCGAAT TGAAAACCGA 1860

CCTAGTGTTG GCGTTCTTCA ATTGCATTTG CTGTGTCTAT TTAGCTCTGG AATCGTAATG 1920

TCCACCTGGT GCTGGACACC TTCTTCAATT GAGACTTGGA AGCGTTATAT AAGGAAAAAG 1980

TGTGGCAAAG AGGTGGTCGA AGAAGTGAAA ATGCCGAAGC ACAAGGTCAT TGCCCAGACA 2040

TGGGCCAAGC GCAAGGATTT CGAGGACAAG GGCAGGCTCT CCATAACGCT CTACAACACC 2100

CACACAGATC CCGTGGGGCT CAACTTCGAT GTGAACGATC TGAACTCTTC TGAGACGAAT 2160

GACATCTCAT CAACTTGGGC TGCATACCTC CCGCAGTGCG TAAAACGTCG CATGGCTTTG 2220

ACGGGAGCAG CGACAGGTAA CTCGTCAAGC CATGGACCGC GAAAAAATTC ATTGGATTCC 2280

GAGATAAGTG TGAGTGTTCG ACATGTTTCC GTTGAATCCC GCAGAAATTC GGTGGACTCG 2340

CAGGTATCAG TGAAAATAGC TGAAATGAAG ACCAAAGTGG CGTCCAGATC AAGGGGAAAA 2400

CACGGAGGCT CTTCCAGCAA CAGAAGAACC CAAAGGAGAA GGGATTATAT AGCAGCTGCC 2460

ACTGGAAAAA GCAGTAGGAG AAGGGAAAGC AGTACTTCAG TGGAGTCGCA GGTCATCGCG 2520

CTCAAGAAAA CGACCTATCC CAATGCTAGT CACAAAGTGG GCGTGTTTGC TCATCACAGC 2580

TCCAAGAAAC AACACAATTA CACCAGCTCC ATGAAGCGAA GGACTGCTAA TGCCGGATTG 2640

GATCCCTCTA TTCTTAATGA ATTCCTGCAG AAAAATGGCG ATTTTATATT CCCATTCCTC 2700

CAAAATCAAG ATATGAGCTC TAGTTCGGAG GAGGATAATT CCAGAGCATC CCAAAAGATT 2760

CAGGATCTTA ACGTGGTTGT AAAGCAGCAG GAAATAAGTG AGGATGATCA CGACGGAATA 2820

AAGATTGAAG AACTGCCAAA TAGCAAACAG GTGGCATTGG AGAACTTTCT TAAAAACATA 2880

AAAAAATCTA ATGAATCCAA TTCTAACCGA CATTCCCGAA ATTCCGCAAG AAGTCAGTCA 2940

AAAAAGTCCC AAAAGAGACA TCTCAAGAAC CCTGCTGCTG ATCTAGATTT CAGGAAGGAC 3000

TGTGTAAAGT ATCGGTCTAA TGACTCACTT AGCTGCTCCT CTGAAGAGCT GGATGTGGCT 3060

TTGGACGTAG GAAGCCTTCT TAACAGCTCT TTTTCTGGAA TATCCATGGG CAAACCACAT 3120

AGTAGAAACA GCAAAACCAG CTGCGATGTG GGCATACAGG CTAATCCTTT CGAGCTAGTT 3180

CCCAGTTACG GAGAAGACGA ACTGCAGCAG GCCATGCGAC TCCTAAACGC AGCCAGCAGA 3240

CAAAGAACTG AAGCAGCCAA TGAGGATTTC GGAGGAACGG AGCTGCAGGG CTTGTTGGGT 3300

CATTCCCATC GGCATCAAAG GGAGCCCACG TTTATGAGCG AGTCGGACAA ACTCAAAATG 3360

TTATTGCTGC CTTCAAAATA GCAAGACTAA ATAAGCAATT GATGCATTTA CTTAAGGTTC 3420

AAAAACTCTT ACAATATTGT AGTTTTTGTT CTAAGAAATC AAATTGTTAG CGCTGAAAAT 3480

AATCGTACAA TCTTATCTAT TTTACGAAAT CGTAATATTG TTATGTTCAC TGTTCAACGA 3540

TTTATAAGAA TATATCGCTT CACTAGAATT GGAAACCCAA ATGATATTTA AAACAAACAA 3600

ATACGAAATT GTAGTACACA AGCCAGAGCA GTTTACATGC GATGAACATT TAGATTCTTC 3660

TTAATCGATT ACTGGAACAG ACTGAGCGAA ACTAGAACTA CGAATTACGA ATACTCATAG 3720

TCATTAGGCT GCAACTTTAT TTTACAGATT CATCACCCCA TCTAGCTTGT AAGCATTCGA 3780

ATCTCTGTGT ACGTTTGTGA ATGACTGTTT CCTTAATCCT GGTACTCACG CCAAAGTAAA 3840

TGCCAAAGAG GATAATAATT TATTTTCATT ATTTTTCTTT GCCGTGGGTA CAGGACTTTA 3900

GATTGTAGAT TATAGATTTA AGTACGATAT AAATAAGCTT CTTGGGCACA CAAATCGTAC 3960

CTCAGAAAGT GCCTTCAAGT TTACAAAATT ATACATAATA ATTTGTGTAA CTAATAAACG 4020

ATTTTAAATC CTCGAGTCT 4039

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2364 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: both
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:

(A) NAME/KEY: CDS
(B) LOCATION: 1..2361

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

ATGGCCGCTG CCCGCCCAGC GCGGGGGCCG GAGCTCCCGC TCCTGGGGCT GCTGCTGCTG 60
CTGCTGCTGG GGGACCCGGG CCGGGGGGCG GCCTCGAGCG GGAACGCGAC CGGGCCTGGG 120
CCTCGGAGCG CGGGCGGGAG CGCGAGGAGG AGCGCGGCGG TGACTGGCCC TCCGCCGCCG 180
CTGAGCCACT GCGGCCGGGC TGCCCCCTGC GAGCCGCTGC GCTACAACGT GTGCCTGGGC 240
TCGGTGCTGC CCTACGGGGC CACCTCCACA CTGCTGGCCG GAGACTCGGA CTCCCAGGAG 300
GAAGCGCACG GCAAGCTCGT GCTCTGGTCG GGCCTCCGGA ATGCCCCCCG CTGCTGGGCA 360
GTGATCCAGC CCCTGCTGTG TGCCGTATAC ATGCCCAAGT GTGAGAATGA CCGGGTGGAG 420
CTGCCCAGCC GTACCCTCTG CCAGGCCACC CGAGGCCCCT GTGCCATCGT GGAGAGGGAG 480
CGGGGCTGGC CTGACTTCCT GCGCTGCACT CCTGACCGCT TCCCTGAAGG CTGCACGAAT 540
GAGGTGCAGA ACATCAAGTT CAACAGTTCA GGCCAGTGCG AAGTGCCCTT GGTTCGGACA 600
GACAACCCCA AGAGCTGGTA CGAGGACGTG GAGGGCTGCG GCATCCAGTG CCAGAACCCG 660
CTCTTCACAG AGGCTGAGCA CCAGGACATG CACAGCTACA TCGCGGCCTT CGGGGCCGTC 720
ACGGGCCTCT GCACGCTCTT CACCCTGGCC ACATTCGTGG CTGACTGGCG GAACTCGAAT 780
CGCTACCCTG CTGTTATTCT CTTCTACGTC AATGCGTGCT TCTTTGTGGG CAGCATTGGC 840
TGGCTGGCCC AGTTCATGGA TGGTGCCCGC CGAGAGATCG TCTGCCGTGC AGATGGCACC 900
ATGAGGCTTG GGGAGCCCAC CTCCAATGAG ACTCTGTCCT GCGTCATCAT CTTTGTCATC 960
GTGTACTACG CCCTGATGGC TGGTGTGGTT TGGTTTGTGG TCCTCACCTA TGCCTGGCAC 1020
ACTTCCTTCA AAGCCCTGGG CACCACCTAC CAGCCTCTCT CGGGCAAGAC CTCCTACTTC 1080
CACCTGCTCA CCTGGTCACT CCCCTTTGTC CTCACTGTGG CAATCCTTGC TGTGGCGCAG 1140
GTGGATGGGG ACTCTGTGAG TGGCATTTGT TTTGTGGGCT ACAAGAACTA CCGATACCGT 1200
GCGGGCTTCG TGCTGGCCCC AATCGGCCTG GTGCTCATCG TGGGAGGCTA CTTCCTCATC 1260
CGAGGAGTCA TGACTCTGTT CTCCATCAAG AGCAACCACC CCGGGCTGCT GAGTGAGAAG 1320
GCTGCCAGCA AGATCAACGA GACCATGCTG CGCCTGGGCA TTTTTGGCTT CCTGGCCTTT 1380
GGCTTTGTGC TCATTACCTT CAGCTGCCAC TTCTACGACT TCTTCAACCA GGCTGAGTGG 1440
GAGCGCAGCT TCCGGGACTA TGTGCTATGT CAGGCCAATG TGACCATCGG GCTGCCCACC 1500
AAGCAGCCCA TCCCTGACTG TGAGATCAAG AATCGCCCGA GCCTTCTGGT GGAGAAGATC 1560
AACCTGTTTG CCATGTTTGG AACTGGCATC GCCATGAGCA CCTGGGTCTG GACCAAGGCC 1620
ACGCTGCTCA TCTGGAGGCG TACCTGGTGC AGGTTGACTG GGCAGAGTGA CGATGAGCCA 1680
AAGCGGATCA AGAAGAGCAA GATGATTGCC AAGGCCTTCT CTAAGCGGCA CGAGCTCCTG 1740
CAGAACCCAG GCCAGGAGCT GTCCTTCAGC ATGCACACTG TGTCCCACGA CGGGCCCGTG 1800

GCGGGCTTGG CCTTTGACCT CAATGAGCCC TCAGCTGATG TCTCCTCTGC CTGGGCCCAG 1860

CATGTCACCA AGATGGTGGC TCGGAGAGGA GCCATACTGC CCCAGGATAT TTCTGTCACC 1920

CCTGTGGCAA CTCCAGTGCC CCCAGAGGAA CAAGCCAACC TGTGGCTGGT TGAGGCAGAG 1980

ATCTCCCCAG AGCTGCAGAA GCGCCTGGGC CGGAAGAAGA AGAGGAGGAA GAGGAAGAAG 2040

GAGGTGTGCC CGCTGGCGCC GCCCCCTGAG CTTCACCCCC CTGCCCCTGC CCCCAGTACC 2100

ATTCCTCGAC TGCCTCAGCT GCCCCGGCAG AAATGCCTGG TGGCTGCAGG TGCCTGGGGA 2160

GCTGGGGACT CTTGCCGACA GGGAGCGTGG ACCCTGGTCT CCAACCCATT CTGCCCAGAG 2220

CCCAGTCCCC CTCAGGATCC ATTTCTGCCC AGTGCACCGG CCCCCGTGGC ATGGGCTCAT 2280

GGCCGCCGAC AGGGCCTGGG GCCTATTCAC TCCCGCACCA ACCTGATGGA CACAGAACTC 2340

ATGGATGCAG ACTCGGACTT CTGA 2364

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2382 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: both
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1..2379

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

ATGGCTGCTG GCCGCCCCGT GCGTGGGCCC GAGCTGGCGC CCCGGAGGCT GCTGCAGTTG 60

CTGCTGCTGG TACTGCTTGG GGGCCGGGGC CGGGGGGCGG CCTTGAGCGG GAACGTGACC 120

GGGCCTGGGC CTCGCAGTGC CGGCGGGAGC GCGAGGAGGA ACGCGCCGGT GACCAGCCCT 180

CCGCCGCCGC TGCTGAGCCA CTGCGGCCGG GCCGCCCACT GCGAGCCTTT GCGCTACAAC 240

GTGTGCCTGG GCTCCGCGCT GCCCTACGGA GCCACCACCA CGCTGCTGGC TGGGGACTCG 300

GACTCGCAGG AGGAAGCGCA CAGCAAGCTC GTGCTCTGGT CCGGCCTCCG GAATGCTCCC 360

CGATGCTGGG CAGTGATCCA GCCCCTGCTG TGTGCTGTCT ACATGCCCAA GTGTGAAAAT 420

GACCGAGTGG AGTTGCCCAG CCGTACCCTC TGCCAGGCCA CCCGAGGCCC CTGTGCCATT 480

GTGGAGCGGG AACGAGGGTG GCCTGACTTT CTGCGTTGCA CGCCGGACCA CTTCCCTGAA 540

GGCTGTCCAA ACGAGGTACA AAACATCAAG TTCAACAGTT CAGGCCAATG TGAAGCACCC 600

TTGGTGAGGA CAGACAACCC CAAGAGCTGG TACGAGGACG TGGAGGGCTG TGGGATCCAG 660

TGCCAGAACC CGCTGTTCAC CGAGGCTGAG CACCAGGACA TGCACAGTTA CATCGCAGCC 720

TTCGGGGCGG TCACCGGCCT CTGTACACTC TTCACCCTGG CCACCTTTGT GGCTGACTGG 780

CGGAACTCCA ATCGCTACCC TGCGGTTATT CTCTTCTATG TCAATGCGTG TTTCTTTGTG 840

GGCAGCATTG GCTGGCTGGC CCAGTTCATG GATGGTGCCC GCCGGGAGAT TGTTTGCCGA 900

GCAGATGGCA CCATGAGATT TGGGGAGCCC ACCTCCAGCG AGACCCTATC CTGTGTCATC 960

ATCTTTGTCA TCGTGTACTA TGCCTTGATG GCTGGAGTAG TGTGGTTCGT GGTCCTCACC 1020

TATGCCTGGC ACACCTCCTT CAAAGCCCTG GGCACCACTT ACCAGCCTCT CTCGGGCAAG 1080

ACATCCTATT TCCACCTGCT CACGTGGTCA CTCCCCTTCG TCCTCACTGT GGCAATCCTT 1140

GCTGTGGCTC AGGTAGATGG GGACTCCGTG AGTGGCATCT GCTTTGTAGG CTACAAGAAC 1200

TATCGGTACC GTGCTGGCTT TGTACTTGCC CCAATTGGCC TGGTGCTTAT TGTGGGAGGC 1260

TACTTCCTCA TCCGAGGGGT CATGACTCTG TTCTCCATCA AGAGCAACCA CCCTGGGCTT 1320

CTGAGTGAGA AGGCAGCCAG CAAGATCAAT GAGACCATGC TGCGCCTGGG CATTTTTGGC 1380

TTCCTCGCCT TTGGCTTCGT GCTCATCACC TTCAGCTGCC ACTTCTATGA CTTCTTCAAC 1440

CAGGCTGAGT GGGAGCGTAG CTTCCGGGAC TATGTGCTAT GCCAAGCCAA TGTGACCATT 1500

GGGCTGCCTA CCAAGAAGCC CATTCCTGAT TGTGAGATCA AGAATCGGCC CAGCCTCCTG 1560

GTGGAGAAGA TCAATCTGTT TGCCATGTTT GGCACTGGCA TTGCCATGAG CACCTGGGTC 1620

TGGACCAAGG CCACCCTGCT CATCTGGAGG CGCACCTGGT GCAGGTTGAC TGGGCACAGT 1680

GATGATGAAC CCAAGAGAAT CAAGAAAAGC AAGATGATTG CCAAGGCCTT CTCTAAGCGG 1740

CGTGAACTGC TGCAGAACCC GGGCCAGGAG CTCTCCTTCA GCATGCACAC TGTCTCCCAT 1800

GATGGACCTG TTGCCGGTTT GGCTTTTGAA CTCAATGAAC CCTCAGCTGA TGTCTCCTCT 1860

GCCTGGGCCC AGCACGTCAC CAAGATGGTG GCTCGAAGAG GAGCCATATT ACCCCAGGAT 1920

GTGTCTGTCA CCCCTGTGGC AACTCCAGTG CCACCAGAAG AACAAGCCAA CCTGTGGCTG 1980

GTTGAGGCAG AGATCTCCCC AGAGTTAGAG AAGCGTTTAG GCCGGAAGAA GAAGCGGAGG 2040

AAGAGGAAGA AGGAGGTGTG CCCCTTGGGG CCAGCCCCTG AACTTCACCA CTCTGCCCCT 2100

GTTCCTGCCA CCAGTGCAGT TCCTCGGCTG CCTCAGCTGC CTCGGCAGAA GTGCCTAGTA 2160

GCTGCAAATG CCTGGGGAAC AGGAGAGCCC TGCCGACAGG GAGCCTGGAC TGTAGTCTCC 2220

AACCCCTTCT GCCCAGAGCC TAGTCCCCAT CAAGATCCAT TTCTCCCTGG TGCCTCAGCC 2280

CCCAGGGTCT GGGCTCAGGG CCGCCTCCAG GGGCTGGGAT CCATTCATTC CCGCACTAAC 2340

CTAATGGAGG CTGAGCTCTT GGATGCAGAC TCGGACTTCT GA 2382

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 3256 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: both
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1..2457

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

GTGGATCCAT TGTGCTCTTG CTTGAATTCG ACTAGCGGGC CCGGGGCCGT CGCCGGGCTC 60

GGGGGCTCCG TGCGGGCCGT GGGGCGGGGG CGGCCCGTGC TGGCTGTGGG CGTGGCGCTG 120

GGCCTGGCGC TCGGCCCCGC GCTGCCCGCG GCGCGCGTCA ACGCGGCATT CGGGTCCTCC 180

CGANNNNCGG NNNGCGGCCT GCGAGCCGTC GCTTCGCTCC TGCCTGGCGT CCCGCTGCCC 240

TGCGCGCACA CCTCCACGCT GCTGGCCGGC GCAACTCGGG ATNGCAGGAG GAGGNGNANG 300

GAAAGCTGCT GCTGTGGTCC GGCCTGCAAT GNGNNNNTGC TGGTACGTGA TCCAGCCGNT 360

GCTGTGTGCT GTCTACATGC CCAAGTGNNN GGATGGGCAG GTGGAGCTGC CAGTCAGNCC 420

CTGTGCCAGG CCACACGTGC ACCCTGCGCC ATCGTGGAGC GCGACGGCTG GCCTGACTTC 480

CTCAAGTGCA CTCCTGACCG CTTCCCCGAG GGCTGCCCGA ACGAGGTGCA GAACATCAAG 540

TTCAACAGCT CAGGGCAGTG CGAGCGCCGT TGGTGCGCAC GTACAACCCC AAGAGCTGGT 600

ATGAGGATGT GGAGGGTTGG TGGAATCCAG NNNAAGAACC CACTCTTCAC TGAGACAGAG 660

CACCGTGAGA TGCAAGTCTA CATCGCCCTT CAGCTCCGTC ACCATCCTTC CTGCACCTTC 720

TTCACTCTGG CCACCTTCGT GGCTGACTGG AGGAACTCCA ACGCTACCCC CGCTGTCATC 780

CTCTTCTATG TCAACGCCTG CTTCTTTGTG GGCAGTATTG GCTGCGTGGC GCAGTTCATG 840

GACGGCGCCC GAGATGAGAT CGTGTGCCGT GCTGATGGCA CCATGAGGCT GGGGGAGCCC 900

ACCTCCAACG AGACGCTCTC CTGCGTCATC ATCTTTGTCA TTGTCTACTA CTCTCTGATG 960

TCGGGCGTCA TCTGGTTTGT CATGCTGACC TACGCCTGGC ACACGTCCTT CAAGGCGCTG 1020

GGCACCACCT ACCAGCCGCT GCTGGGCAAG ACCTCCTACT TCCACCTCAT CACCTGGTCC 1080

ATCCCTTTCG TACTCACCGT GGCCATCCTG GCTGTGGCAC AGGTGGATGG TGACTCCGTC 1140

AGCGGTATCT GCTTCGTGGG TTACAAGAAC TATCGCTACC GTGCCGGCTT TGTCCTGGCA 1200

CCCATCGGGC TCGTCCTCAT CGTTGGGGGC TATTTCCTCA TTCGGGGGGT CATGACGCTC 1260

TTCTCCATCA AGAGCAACCA CCCCGGGCTG CTGAGTGAGA AGGCGGCCAG CAAGATCAAC 1320

GAAACCATGC TGCGGCTGGG CATCTTTGGG TTCTTGGCCT TTGGCTTTGT CTTCATCACT 1380

TTTGGCTGCC ACTTCTACGA CTTCTTCAAC CAGGCGGAGT GGGAGCGAAG CTTTCGGGAA 1440

TATGTCCTGT GTGAGGCCAA CGTGACCATC GCTACGCAGA CCAATAAACC CATCCCGGAG 1500

TGTGAGATTA AGAACCGGCC GAGCCTGCTG GTGGAGAAGA TCAACCTCTT TGCCATGTTT 1560

GGCACTGGCA TCTCCATGAG CACCTGGGTC TGGACCAAGG CCACCCTGCT CATCTGGAAG 1620

CGCACCTGGT GCAGGCTGAC AGGGCAGAGC GACGACCAGC CCAAGAGGAT CAAGAAGAGC 1680

AAGATGATTG CCAAAGCCTT CTCCAAGCGC AGGGAGCTTC TGCGTGACCC GGGCCGGGAG 1740

CTGTCCTTCA GTATGCACAC CGTCTCGCAC GATGGCCCCG TGGCTGGTTT GGCGTTTGAC 1800

ATCAATGAGC CATCAGCCGA TGTGTCCTCC GCGTGGGCTC AGCACGTCAC CAAGATGGTG 1860

GCCAGGAGAG GGGCTATCCT GCCCCAGGAT GTCTCCGTCA CGCCGGTGGC AACACCTGTG 1920

CCACCGGAGG AGCGGAGCAA CCTCTGGGTG GTGGAGGCCG ATGTCTCCCC AGAGCTGCAG 1980

AAGCGCAGCC GCAAGAAGAA GCGGAGGAAG AAGAAGAAGG AGGAGGTGTG CCCCGAGCGC 2040

CGCGCCGGGC TCTCCGTGGC CCCCCTGACC CCCAGCTCCG TGCCTCGCCT GCCTCGGCTG 2100

CCCCAGCAGC CCTGCTTGGT GGCCATCCCC CGGCATAGAG GGGACACCTT CATCCCCACT 2160

GTCCTCCCGG GGCTGTCCAA CGGTGCTGGG GGGCTGTGGG ACGGCCGGCG CCGAGCCCAC 2220

GTCCCCCACT TCATCACCAA CCCCTTCTGC CCTGAGAGTG GCTCCCCAGA GGATGAGGAG 2280

AACCCCGGCC CCAGCGTCGG GCACCGGCAG CACAACGGGG GCCYTCGATG GCCACCTGAG 2340

CCCCTTCCTG GTGGCAGTGG GGTGACGAGG ACTCGGGGCA GACGCGCCGG CTTGGCTCCC 2400

ATCCACTCCC GGACCAACCT GGTGAACGCG GAGCTGCTGG ACGCCGACTT AGACTTCTGA 2460

GCCCTGCAGG ATCCTGGGGA CAACGGAGCC CACCGGCATC TGGGTAGCCC CAAGGGACGC 2520

TGGAGCCCAC CCAACACCGG GGTCCGGATG GATTTGATGT TCATCCCAAC CCACCAACGG 2580

GATTTGAGGA TGGGAGGAGA GAAGAAACTG GTGGGGCAGC ACCCCTGGAA GGCGCTGCCA 2640

GGATTTAGGG GTGAAGGGGA CGCTCCCTCA CACCCAGCGC TGGTAGGCAG CTTAAGGTGT 2700

TGATTTCGGT CCCCACACAT GGACTGCTCC GCCGCCCCAC AGCTAGATGG TACGTAGAGC 2760

TTCCCAACAC TTTTACGGTG CCAATAGGGT TTTTAAACAG TTCTTTTTGT ATTCTTTGTG 2820

ATACACCGAG ACGTGGCCGC CCTGCACGGG GTGCAGCAGC ATCCCCGTTT TTCGGGCTCT 2880

GCTGCGGGGA TCCCAAAGTG CCTTCCAGCT CCCCTCGGCC GTCTGAGCGC ACCTGAGAAA 2940

AGCTTTGGTT TTTGTTCGTT TTTAAATCTG TTTTTTAAAG AAAAAGGAAC AAATTATATC 3000

CGAGCCCTGA CGTAGGAGGA CACCTGTCCT TGCTGGTGCT TTGTATCTGC CCCTTAGCCC 3060

TGTAAATGTC TTTTGAGTGT TTATTAAACC CCGGTAGGCT CACGGCTTCC TCCTAACCCC 3120

CCCCCTACCC CCCACCTCCC CCCTCCCGGC GACCTCCACG GGTTGATTTT TGCTGTGTAA 3180

AGCAGAAACC TTCCTGTATC AGTATTAAAT TTGCCAAGTT TCCAATTGCA AAAAAAAAAA 3240

AAAAAAAAAA ACTCGA 3256

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1036 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

Met Gln Tyr Leu Asn Phe Pro Arg Met Pro Asn Ile Met Met Phe Leu
1 5 10 15

Glu Val Ala Ile Leu Cys Leu Trp Val Val Ala Asp Ala Ser Ala Ser
20 25 30

Ser Ala Lys Phe Gly Ser Thr Thr Pro Ala Ser Ala Gln Gln Ser Asp
35 40 45

Val Glu Leu Glu Pro Ile Asn Gly Thr Leu Asn Tyr Arg Leu Tyr Ala
50 55 60

Lys Lys Gly Arg Asp Asp Lys Pro Trp Phe Asp Gly Leu Asp Ser Arg
65 70 75 80

His Ile Gln Cys Val Arg Arg Ala Arg Cys Tyr Pro Thr Ser Asn Ala
85 90 95

Thr Asn Thr Cys Phe Gly Ser Lys Leu Pro Tyr Glu Leu Ser Ser Leu
100 105 110

Asp Leu Thr Asp Phe His Thr Glu Lys Glu Leu Asn Asp Lys Leu Asn
115 120 125

Asp Tyr Tyr Ala Leu Lys His Val Pro Lys Cys Trp Ala Ala Ile Gln
130 135 140

Pro Phe Leu Cys Ala Val Phe Lys Pro Lys Cys Glu Lys Ile Asn Gly
145 150 155 160

Glu Asp Met Val Tyr Leu Pro Ser Tyr Glu Met Cys Arg Ile Thr Met
165 170 175

Glu Pro Cys Arg Ile Leu Tyr Asn Thr Thr Phe Phe Pro Lys Phe Leu
180 185 190

Arg Cys Asn Glu Thr Leu Phe Pro Thr Lys Cys Thr Asn Gly Ala Arg
195 200 205

Gly Met Lys Phe Asn Gly Thr Gly Gln Cys Leu Ser Pro Leu Val Pro
210 215 220

Thr Asp Thr Ser Ala Ser Tyr Tyr Pro Gly Ile Glu Gly Cys Gly Val
225 230 235 240

Arg Cys Lys Asp Pro Leu Tyr Thr Asp Asp Glu His Arg Gln Ile His
245 250 255

Lys Leu Ile Gly Trp Ala Gly Ser Ile Cys Leu Leu Ser Asn Leu Phe
260 265 270

Val Val Ser Thr Phe Phe Ile Asp Trp Lys Asn Ala Asn Lys Tyr Pro
275 280 285

Ala Val Ile Val Phe Tyr Ile Asn Leu Cys Phe Leu Ile Ala Cys Val
290 295 300

Gly Trp Leu Leu Gln Phe Thr Ser Gly Ser Arg Glu Asp Ile Val Cys
305 310 315 320

Arg Lys Asp Gly Thr Leu Arg His Ser Glu Pro Thr Ala Gly Glu Asn
325 330 335

Leu Ser Cys Ile Val Ile Phe Val Leu Val Tyr Tyr Phe Leu Thr Ala
340 345 350

Gly Met Val Trp Phe Val Phe Leu Thr Tyr Ala Trp His Trp Arg Ala
355 360 365

Met Gly His Val Gln Asp Arg Ile Asp Lys Lys Gly Ser Tyr Phe His
370 375 380

Leu Val Ala Trp Ser Leu Pro Leu Val Leu Thr Ile Thr Thr Met Ala
385 390 395 400

Phe Ser Glu Val Asp Gly Asn Ser Ile Val Gly Ile Cys Phe Val Gly
405 410 415

Tyr Ile Asn His Ser Met Arg Ala Gly Leu Leu Leu Gly Pro Leu Cys
420 425 430

Gly Val Ile Leu Ile Gly Gly Tyr Phe Ile Thr Arg Gly Met Val Met
435 440 445

Leu Phe Gly Leu Lys His Phe Ala Asn Asp Ile Lys Ser Thr Ser Ala
450 455 460

Ser Asn Lys Ile His Leu Ile Ile Met Arg Met Gly Val Cys Ala Leu
465 470 475 480

Leu Thr Leu Val Phe Ile Leu Val Ala Ile Ala Cys His Val Thr Glu
485 490 495

Phe Arg His Ala Asp Glu Trp Ala Gln Ser Phe Arg Gln Phe Ile Ile
500 505 510

Cys Lys Ile Ser Ser Val Phe Glu Glu Lys Ser Ser Cys Arg Ile Glu
515 520 525

Asn Arg Pro Ser Val Gly Val Leu Gln Leu His Leu Leu Cys Leu Phe
530 535 540

Ser Ser Gly Ile Val Met Ser Thr Trp Cys Trp Thr Pro Ser Ser Ile
545 550 555 560

Glu Thr Trp Lys Arg Tyr Ile Arg Lys Lys Cys Gly Lys Glu Val Val
565 570 575

Glu Glu Val Lys Met Pro Lys His Lys Val Ile Ala Gln Thr Trp Ala
580 585 590

Lys Arg Lys Asp Phe Glu Asp Lys Gly Arg Leu Ser Ile Thr Leu Tyr
595 600 605

Asn Thr His Thr Asp Pro Val Gly Leu Asn Phe Asp Val Asn Asp Leu
610 615 620

Asn Ser Ser Glu Thr Asn Asp Ile Ser Ser Thr Trp Ala Ala Tyr Leu
625 630 635 640

Pro Gln Cys Val Lys Arg Arg Met Ala Leu Thr Gly Ala Ala Thr Gly
645 650 655

Asn Ser Ser Ser His Gly Pro Arg Lys Asn Ser Leu Asp Ser Glu Ile
660 665 670

Ser Val Ser Val Arg His Val Ser Val Glu Ser Arg Arg Asn Ser Val
675 680 685

Asp Ser Gln Val Ser Val Lys Ile Ala Glu Met Lys Thr Lys Val Ala
690 695 700

Ser Arg Ser Arg Gly Lys His Gly Gly Ser Ser Ser Asn Arg Arg Thr
705 710 715 720

Gln Arg Arg Arg Asp Tyr Ile Ala Ala Ala Thr Gly Lys Ser Ser Arg
725 730 735

```
Arg Arg Glu Ser Ser Thr Ser Val Glu Ser Gln Val Ile Ala Leu Lys
            740                 745                 750

Lys Thr Thr Tyr Pro Asn Ala Ser His Lys Val Gly Val Phe Ala His
        755                 760                 765

His Ser Ser Lys Lys Gln His Asn Tyr Thr Ser Ser Met Lys Arg Arg
    770                 775                 780

Thr Ala Asn Ala Gly Leu Asp Pro Ser Ile Leu Asn Glu Phe Leu Gln
785                 790                 795                 800

Lys Asn Gly Asp Phe Ile Phe Pro Phe Leu Gln Asn Gln Asp Met Ser
                805                 810                 815

Ser Ser Ser Glu Glu Asp Asn Ser Arg Ala Ser Gln Lys Ile Gln Asp
            820                 825                 830

Leu Asn Val Val Val Lys Gln Gln Glu Ile Ser Glu Asp Asp His Asp
        835                 840                 845

Gly Ile Lys Ile Glu Glu Leu Pro Asn Ser Lys Gln Val Ala Leu Glu
    850                 855                 860

Asn Phe Leu Lys Asn Ile Lys Lys Ser Asn Glu Ser Asn Ser Asn Arg
865                 870                 875                 880

His Ser Arg Asn Ser Ala Arg Ser Gln Ser Lys Lys Ser Gln Lys Arg
                885                 890                 895

His Leu Lys Asn Pro Ala Ala Asp Leu Asp Phe Arg Lys Asp Cys Val
            900                 905                 910

Lys Tyr Arg Ser Asn Asp Ser Leu Ser Cys Ser Ser Glu Glu Leu Asp
        915                 920                 925

Val Ala Leu Asp Val Gly Ser Leu Leu Asn Ser Ser Phe Ser Gly Ile
    930                 935                 940

Ser Met Gly Lys Pro His Ser Arg Asn Ser Lys Thr Ser Cys Asp Val
945                 950                 955                 960

Gly Ile Gln Ala Asn Pro Phe Glu Leu Val Pro Ser Tyr Gly Glu Asp
                965                 970                 975

Glu Leu Gln Gln Ala Met Arg Leu Leu Asn Ala Ala Ser Arg Gln Arg
            980                 985                 990

Thr Glu Ala Ala Asn Glu Asp Phe Gly Gly Thr Glu Leu Gln Gly Leu
        995                 1000                1005

Leu Gly His Ser His Arg His Gln Arg Glu Pro Thr Phe Met Ser Glu
    1010                1015                1020

Ser Asp Lys Leu Lys Met Leu Leu Leu Pro Ser Lys
1025                1030                1035
```

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 787 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

```
Met Ala Ala Ala Arg Pro Ala Arg Gly Pro Glu Leu Pro Leu Leu Gly
 1               5                  10                  15

Leu Leu Leu Leu Leu Leu Leu Gly Asp Pro Gly Arg Gly Ala Ala Ser
            20                  25                  30

Ser Gly Asn Ala Thr Gly Pro Gly Pro Arg Ser Ala Gly Gly Ser Ala
        35                  40                  45

Arg Arg Ser Ala Ala Val Thr Gly Pro Pro Pro Pro Leu Ser His Cys
    50                  55                  60

Gly Arg Ala Ala Pro Cys Glu Pro Leu Arg Tyr Asn Val Cys Leu Gly
65                  70                  75                  80

Ser Val Leu Pro Tyr Gly Ala Thr Ser Thr Leu Leu Ala Gly Asp Ser
                85                  90                  95

Asp Ser Gln Glu Glu Ala His Gly Lys Leu Val Leu Trp Ser Gly Leu
            100                 105                 110

Arg Asn Ala Pro Arg Cys Trp Ala Val Ile Gln Pro Leu Leu Cys Ala
        115                 120                 125

Val Tyr Met Pro Lys Cys Glu Asn Asp Arg Val Glu Leu Pro Ser Arg
    130                 135                 140

Thr Leu Cys Gln Ala Thr Arg Gly Pro Cys Ala Ile Val Glu Arg Glu
145                 150                 155                 160

Arg Gly Trp Pro Asp Phe Leu Arg Cys Thr Pro Asp Arg Phe Pro Glu
                165                 170                 175

Gly Cys Thr Asn Glu Val Gln Asn Ile Lys Phe Asn Ser Ser Gly Gln
            180                 185                 190

Cys Glu Val Pro Leu Val Arg Thr Asp Asn Pro Lys Ser Trp Tyr Glu
        195                 200                 205

Asp Val Glu Gly Cys Gly Ile Gln Cys Gln Asn Pro Leu Phe Thr Glu
    210                 215                 220

Ala Glu His Gln Asp Met His Ser Tyr Ile Ala Ala Phe Gly Ala Val
225                 230                 235                 240

Thr Gly Leu Cys Thr Leu Phe Thr Leu Ala Thr Phe Val Ala Asp Trp
                245                 250                 255

Arg Asn Ser Asn Arg Tyr Pro Ala Val Ile Leu Phe Tyr Val Asn Ala
            260                 265                 270

Cys Phe Phe Val Gly Ser Ile Gly Trp Leu Ala Gln Phe Met Asp Gly
        275                 280                 285

Ala Arg Arg Glu Ile Val Cys Arg Ala Asp Gly Thr Met Arg Leu Gly
    290                 295                 300

Glu Pro Thr Ser Asn Glu Thr Leu Ser Cys Val Ile Ile Phe Val Ile
305                 310                 315                 320

Val Tyr Tyr Ala Leu Met Ala Gly Val Val Trp Phe Val Val Leu Thr
                325                 330                 335
```

```
Tyr Ala Trp His Thr Ser Phe Lys Ala Leu Gly Thr Thr Tyr Gln Pro
            340                 345                 350

Leu Ser Gly Lys Thr Ser Tyr Phe His Leu Leu Thr Trp Ser Leu Pro
        355                 360                 365

Phe Val Leu Thr Val Ala Ile Leu Ala Val Ala Gln Val Asp Gly Asp
    370                 375                 380

Ser Val Ser Gly Ile Cys Phe Val Gly Tyr Lys Asn Tyr Arg Tyr Arg
385                 390                 395                 400

Ala Gly Phe Val Leu Ala Pro Ile Gly Leu Val Leu Ile Val Gly Gly
                405                 410                 415

Tyr Phe Leu Ile Arg Gly Val Met Thr Leu Phe Ser Ile Lys Ser Asn
            420                 425                 430

His Pro Gly Leu Leu Ser Glu Lys Ala Ala Ser Lys Ile Asn Glu Thr
        435                 440                 445

Met Leu Arg Leu Gly Ile Phe Gly Phe Leu Ala Phe Gly Phe Val Leu
    450                 455                 460

Ile Thr Phe Ser Cys His Phe Tyr Asp Phe Phe Asn Gln Ala Glu Trp
465                 470                 475                 480

Glu Arg Ser Phe Arg Asp Tyr Val Leu Cys Gln Ala Asn Val Thr Ile
                485                 490                 495

Gly Leu Pro Thr Lys Gln Pro Ile Pro Asp Cys Glu Ile Lys Asn Arg
            500                 505                 510

Pro Ser Leu Leu Val Glu Lys Ile Asn Leu Phe Ala Met Phe Gly Thr
        515                 520                 525

Gly Ile Ala Met Ser Thr Trp Val Trp Thr Lys Ala Thr Leu Leu Ile
    530                 535                 540

Trp Arg Arg Thr Trp Cys Arg Leu Thr Gly Gln Ser Asp Asp Glu Pro
545                 550                 555                 560

Lys Arg Ile Lys Lys Ser Lys Met Ile Ala Lys Ala Phe Ser Lys Arg
                565                 570                 575

His Glu Leu Leu Gln Asn Pro Gly Gln Glu Leu Ser Phe Ser Met His
            580                 585                 590

Thr Val Ser His Asp Gly Pro Val Ala Gly Leu Ala Phe Asp Leu Asn
        595                 600                 605

Glu Pro Ser Ala Asp Val Ser Ser Ala Trp Ala Gln His Val Thr Lys
    610                 615                 620

Met Val Ala Arg Arg Gly Ala Ile Leu Pro Gln Asp Ile Ser Val Thr
625                 630                 635                 640

Pro Val Ala Thr Pro Val Pro Pro Glu Glu Gln Ala Asn Leu Trp Leu
                645                 650                 655

Val Glu Ala Glu Ile Ser Pro Glu Leu Gln Lys Arg Leu Gly Arg Lys
            660                 665                 670

Lys Lys Arg Arg Lys Arg Lys Lys Glu Val Cys Pro Leu Ala Pro Pro
        675                 680                 685
```

```
Pro Glu Leu His Pro Pro Ala Pro Ala Pro Ser Thr Ile Pro Arg Leu
    690                 695                 700

Pro Gln Leu Pro Arg Gln Lys Cys Leu Val Ala Ala Gly Ala Trp Gly
705                 710                 715                 720

Ala Gly Asp Ser Cys Arg Gln Gly Ala Trp Thr Leu Val Ser Asn Pro
                725                 730                 735

Phe Cys Pro Glu Pro Ser Pro Pro Gln Asp Pro Phe Leu Pro Ser Ala
            740                 745                 750

Pro Ala Pro Val Ala Trp Ala His Gly Arg Arg Gln Gly Leu Gly Pro
        755                 760                 765

Ile His Ser Arg Thr Asn Leu Met Asp Thr Glu Leu Met Asp Ala Asp
    770                 775                 780

Ser Asp Phe
785
```

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 793 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

```
Met Ala Ala Gly Arg Pro Val Arg Gly Pro Glu Leu Ala Pro Arg Arg
  1               5                  10                  15

Leu Leu Gln Leu Leu Leu Leu Val Leu Leu Gly Gly Arg Gly Arg Gly
            20                  25                  30

Ala Ala Leu Ser Gly Asn Val Thr Gly Pro Gly Pro Arg Ser Ala Gly
        35                  40                  45

Gly Ser Ala Arg Arg Asn Ala Pro Val Thr Ser Pro Pro Pro Pro Leu
    50                  55                  60

Leu Ser His Cys Gly Arg Ala Ala His Cys Glu Pro Leu Arg Tyr Asn
65                  70                  75                  80

Val Cys Leu Gly Ser Ala Leu Pro Tyr Gly Ala Thr Thr Thr Leu Leu
                85                  90                  95

Ala Gly Asp Ser Asp Ser Gln Glu Glu Ala His Ser Lys Leu Val Leu
            100                 105                 110

Trp Ser Gly Leu Arg Asn Ala Pro Arg Cys Trp Ala Val Ile Gln Pro
        115                 120                 125

Leu Leu Cys Ala Val Tyr Met Pro Lys Cys Glu Asn Asp Arg Val Glu
    130                 135                 140

Leu Pro Ser Arg Thr Leu Cys Gln Ala Thr Arg Gly Pro Cys Ala Ile
145                 150                 155                 160

Val Glu Arg Glu Arg Gly Trp Pro Asp Phe Leu Arg Cys Thr Pro Asp
                165                 170                 175
```

```
His Phe Pro Glu Gly Cys Pro Asn Glu Val Gln Asn Ile Lys Phe Asn
            180                 185                 190

Ser Ser Gly Gln Cys Glu Ala Pro Leu Val Arg Thr Asp Asn Pro Lys
        195                 200                 205

Ser Trp Tyr Glu Asp Val Glu Gly Cys Gly Ile Gln Cys Gln Asn Pro
    210                 215                 220

Leu Phe Thr Glu Ala Glu His Gln Asp Met His Ser Tyr Ile Ala Ala
225                 230                 235                 240

Phe Gly Ala Val Thr Gly Leu Cys Thr Leu Phe Thr Leu Ala Thr Phe
                245                 250                 255

Val Ala Asp Trp Arg Asn Ser Asn Arg Tyr Pro Ala Val Ile Leu Phe
            260                 265                 270

Tyr Val Asn Ala Cys Phe Phe Val Gly Ser Ile Gly Trp Leu Ala Gln
        275                 280                 285

Phe Met Asp Gly Ala Arg Arg Glu Ile Val Cys Arg Ala Asp Gly Thr
    290                 295                 300

Met Arg Phe Gly Glu Pro Thr Ser Ser Glu Thr Leu Ser Cys Val Ile
305                 310                 315                 320

Ile Phe Val Ile Val Tyr Tyr Ala Leu Met Ala Gly Val Val Trp Phe
                325                 330                 335

Val Val Leu Thr Tyr Ala Trp His Thr Ser Phe Lys Ala Leu Gly Thr
            340                 345                 350

Thr Tyr Gln Pro Leu Ser Gly Lys Thr Ser Tyr Phe His Leu Leu Thr
        355                 360                 365

Trp Ser Leu Pro Phe Val Leu Thr Val Ala Ile Leu Ala Val Ala Gln
    370                 375                 380

Val Asp Gly Asp Ser Val Ser Gly Ile Cys Phe Val Gly Tyr Lys Asn
385                 390                 395                 400

Tyr Arg Tyr Arg Ala Gly Phe Val Leu Ala Pro Ile Gly Leu Val Leu
                405                 410                 415

Ile Val Gly Gly Tyr Phe Leu Ile Arg Gly Val Met Thr Leu Phe Ser
            420                 425                 430

Ile Lys Ser Asn His Pro Gly Leu Leu Ser Glu Lys Ala Ala Ser Lys
        435                 440                 445

Ile Asn Glu Thr Met Leu Arg Leu Gly Ile Phe Gly Phe Leu Ala Phe
    450                 455                 460

Gly Phe Val Leu Ile Thr Phe Ser Cys His Phe Tyr Asp Phe Phe Asn
465                 470                 475                 480

Gln Ala Glu Trp Glu Arg Ser Phe Arg Asp Tyr Val Leu Cys Gln Ala
                485                 490                 495

Asn Val Thr Ile Gly Leu Pro Thr Lys Lys Pro Ile Pro Asp Cys Glu
            500                 505                 510

Ile Lys Asn Arg Pro Ser Leu Leu Val Glu Lys Ile Asn Leu Phe Ala
        515                 520                 525
```

```
Met Phe Gly Thr Gly Ile Ala Met Ser Thr Trp Val Trp Thr Lys Ala
    530                 535                 540

Thr Leu Leu Ile Trp Arg Arg Thr Trp Cys Arg Leu Thr Gly His Ser
545                 550                 555                 560

Asp Asp Glu Pro Lys Arg Ile Lys Lys Ser Lys Met Ile Ala Lys Ala
                565                 570                 575

Phe Ser Lys Arg Arg Glu Leu Leu Gln Asn Pro Gly Gln Glu Leu Ser
            580                 585                 590

Phe Ser Met His Thr Val Ser His Asp Gly Pro Val Ala Gly Leu Ala
        595                 600                 605

Phe Glu Leu Asn Glu Pro Ser Ala Asp Val Ser Ser Ala Trp Ala Gln
    610                 615                 620

His Val Thr Lys Met Val Ala Arg Arg Gly Ala Ile Leu Pro Gln Asp
625                 630                 635                 640

Val Ser Val Thr Pro Val Ala Thr Pro Val Pro Pro Glu Glu Gln Ala
                645                 650                 655

Asn Leu Trp Leu Val Glu Ala Glu Ile Ser Pro Glu Leu Glu Lys Arg
            660                 665                 670

Leu Gly Arg Lys Lys Lys Arg Arg Lys Arg Lys Lys Glu Val Cys Pro
        675                 680                 685

Leu Gly Pro Ala Pro Glu Leu His His Ser Ala Pro Val Pro Ala Thr
    690                 695                 700

Ser Ala Val Pro Arg Leu Pro Gln Leu Pro Arg Gln Lys Cys Leu Val
705                 710                 715                 720

Ala Ala Asn Ala Trp Gly Thr Gly Glu Pro Cys Arg Gln Gly Ala Trp
                725                 730                 735

Thr Val Val Ser Asn Pro Phe Cys Pro Glu Pro Ser Pro His Gln Asp
            740                 745                 750

Pro Phe Leu Pro Gly Ala Ser Ala Pro Arg Val Trp Ala Gln Gly Arg
        755                 760                 765

Leu Gln Gly Leu Gly Ser Ile His Ser Arg Thr Asn Leu Met Glu Ala
    770                 775                 780

Glu Leu Leu Asp Ala Asp Ser Asp Phe
785                 790
```

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 819 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

```
Val Asp Pro Leu Cys Ser Cys Leu Asn Ser Thr Ser Gly Pro Gly Ala
```

```
1               5                   10                  15

Val Ala Gly Leu Gly Gly Ser Val Arg Ala Val Gly Arg Gly Arg Pro
            20                  25                  30

Val Leu Ala Val Gly Val Ala Leu Gly Leu Ala Leu Gly Pro Ala Leu
        35                  40                  45

Pro Ala Ala Arg Val Asn Ala Ala Phe Gly Ser Ser Arg Xaa Xaa Xaa
    50                  55                  60

Xaa Gly Leu Arg Ala Val Ala Ser Leu Leu Pro Gly Val Pro Leu Pro
65                  70                  75                  80

Cys Ala His Thr Ser Thr Leu Leu Ala Gly Ala Thr Arg Asp Xaa Arg
                85                  90                  95

Arg Arg Xaa Xaa Glu Ser Cys Cys Cys Gly Pro Ala Cys Asn Xaa Xaa
            100                 105                 110

Xaa Leu Val Arg Asp Pro Ala Xaa Ala Val Cys Cys Leu His Ala Gln
        115                 120                 125

Val Xaa Gly Trp Ala Gly Gly Ala Ala Ser Gln Xaa Leu Cys Gln Ala
    130                 135                 140

Thr Arg Ala Pro Cys Ala Ile Val Glu Arg Asp Gly Trp Pro Asp Phe
145                 150                 155                 160

Leu Lys Cys Thr Pro Asp Arg Phe Pro Glu Gly Cys Pro Asn Glu Val
                165                 170                 175

Gln Asn Ile Lys Phe Asn Ser Ser Gly Gln Cys Glu Arg Arg Trp Cys
            180                 185                 190

Ala Arg Thr Thr Pro Arg Ala Gly Met Arg Met Trp Arg Val Gly Gly
        195                 200                 205

Ile Gln Xaa Lys Asn Pro Leu Phe Thr Glu Thr Glu His Arg Glu Met
    210                 215                 220

Gln Val Tyr Ile Ala Leu Gln Leu Arg His His Pro Ser Cys Thr Phe
225                 230                 235                 240

Phe Thr Leu Ala Thr Phe Val Ala Asp Trp Arg Asn Ser Asn Ala Thr
                245                 250                 255

Pro Ala Val Ile Leu Phe Tyr Val Asn Ala Cys Phe Phe Val Gly Ser
            260                 265                 270

Ile Gly Cys Val Ala Gln Phe Met Asp Gly Ala Arg Asp Glu Ile Val
        275                 280                 285

Cys Arg Ala Asp Gly Thr Met Arg Leu Gly Glu Pro Thr Ser Asn Glu
    290                 295                 300

Thr Leu Ser Cys Val Ile Ile Phe Val Ile Val Tyr Tyr Ser Leu Met
305                 310                 315                 320

Ser Gly Val Ile Trp Phe Val Met Leu Thr Tyr Ala Trp His Thr Ser
                325                 330                 335

Phe Lys Ala Leu Gly Thr Thr Tyr Gln Pro Leu Leu Gly Lys Thr Ser
            340                 345                 350

Tyr Phe His Leu Ile Thr Trp Ser Ile Pro Phe Val Leu Thr Val Ala
```

```
          355                 360                 365

Ile Leu Ala Val Ala Gln Val Asp Gly Asp Ser Val Ser Gly Ile Cys
    370                 375                 380

Phe Val Gly Tyr Lys Asn Tyr Arg Tyr Arg Ala Gly Phe Val Leu Ala
385                 390                 395                 400

Pro Ile Gly Leu Val Leu Ile Val Gly Gly Tyr Phe Leu Ile Arg Gly
                405                 410                 415

Val Met Thr Leu Phe Ser Ile Lys Ser Asn His Pro Gly Leu Leu Ser
            420                 425                 430

Glu Lys Ala Ala Ser Lys Ile Asn Glu Thr Met Leu Arg Leu Gly Ile
        435                 440                 445

Phe Gly Phe Leu Ala Phe Gly Phe Val Phe Ile Thr Phe Gly Cys His
    450                 455                 460

Phe Tyr Asp Phe Phe Asn Gln Ala Glu Trp Glu Arg Ser Phe Arg Glu
465                 470                 475                 480

Tyr Val Leu Cys Glu Ala Asn Val Thr Ile Ala Thr Gln Thr Asn Lys
                485                 490                 495

Pro Ile Pro Glu Cys Glu Ile Lys Asn Arg Pro Ser Leu Leu Val Glu
            500                 505                 510

Lys Ile Asn Leu Phe Ala Met Phe Gly Thr Gly Ile Ser Met Ser Thr
        515                 520                 525

Trp Val Trp Thr Lys Ala Thr Leu Leu Ile Trp Lys Arg Thr Trp Cys
    530                 535                 540

Arg Leu Thr Gly Gln Ser Asp Asp Gln Pro Lys Arg Ile Lys Lys Ser
545                 550                 555                 560

Lys Met Ile Ala Lys Ala Phe Ser Lys Arg Arg Glu Leu Leu Arg Asp
                565                 570                 575

Pro Gly Arg Glu Leu Ser Phe Ser Met His Thr Val Ser His Asp Gly
            580                 585                 590

Pro Val Ala Gly Leu Ala Phe Asp Ile Asn Glu Pro Ser Ala Asp Val
        595                 600                 605

Ser Ser Ala Trp Ala Gln His Val Thr Lys Met Val Ala Arg Arg Gly
    610                 615                 620

Ala Ile Leu Pro Gln Asp Val Ser Val Thr Pro Val Ala Thr Pro Val
625                 630                 635                 640

Pro Pro Glu Glu Arg Ser Asn Leu Trp Val Val Glu Ala Asp Val Ser
                645                 650                 655

Pro Glu Leu Gln Lys Arg Ser Arg Lys Lys Lys Arg Arg Lys Lys Lys
            660                 665                 670

Lys Glu Glu Val Cys Pro Glu Arg Arg Ala Gly Leu Ser Val Ala Pro
        675                 680                 685

Leu Thr Pro Ser Ser Val Pro Arg Leu Pro Arg Leu Pro Gln Gln Pro
    690                 695                 700

Cys Leu Val Ala Ile Pro Arg His Arg Gly Asp Thr Phe Ile Pro Thr
```

705 710 715 720

Val Leu Pro Gly Leu Ser Asn Gly Ala Gly Gly Leu Trp Asp Gly Arg
725 730 735

Arg Arg Ala His Val Pro His Phe Ile Thr Asn Pro Phe Cys Pro Glu
740 745 750

Ser Gly Ser Pro Glu Asp Glu Glu Asn Pro Gly Pro Ser Val Gly His
755 760 765

Arg Gln His Asn Gly Gly Xaa Arg Trp Pro Pro Glu Pro Leu Pro Gly
770 775 780

Gly Ser Gly Val Thr Arg Thr Arg Gly Arg Arg Ala Gly Leu Ala Pro
785 790 795 800

Ile His Ser Arg Thr Asn Leu Val Asn Ala Glu Leu Leu Asp Ala Asp
805 810 815

Leu Asp Phe

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 4617 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: both
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

GGGTATATTT AAACTGGCCC TGAATGGTCG CACATTTGTT GCTTCAGCTT CACGTTGAGT 60

CGCGTTTATT TTTGTTTTCC TCCTTGGTTC TTGTTTTTTT TTTCTCTTTT TTAAAAACAC 120

TGCAATCCTT CAAAGGGAAC AATTAAACCA GGATATCCCG AGTCCAACGA AAGCGCCTTC 180

ACGATGCACT GACGAACCGC TGAGCAAGAA ACGGGCATTC AGCCCATTAA AATTCACACA 240

AGGCAGCTAT TTAATGATTT TTATAAGCGG ATGCAGTACT TAAACTTTCC GCGCATGCCA 300

AACATTATGT GAGTGACTCG CCTCTGGCAA ATGGTGTGGG CAATTAAAAT AACTTGGTTA 360

TTTACAAACC GCAGGATGTT CCTGGAGGTT GCGATCTTAT GCCTGTGGGT GGTCGCAGAC 420

GCATCGGCCA GTTCGGCCAA GTTCGGCAGC ACAACGCCCG CAAGTGCGCA GCAGTCGGAT 480

GTGGAACTGG AGCCCATCAA TGGGACTCTC AATTACCGAC TGTACGCCAA GAAAGGCAGG 540

GACGACAAAC CCTGGTTTGA TGGCCTAGAC AGCAGGCACA TCCAGTGTGT CCGACGTGCC 600

CGTTGCTACC CCACCTCGAA CGCAACCAAC ACCTGTTTCG GCTCAAAATT GCCCTATGAG 660

CTGAGCAGCC TAGATCTCAC CGACTTCCAC ACCGAAAAGG AGCTGAACGA TAAGCTGAAC 720

GACTACTATG CCCTGAAGCA CGTGCCCAAA TGTTGGGCAG CTATACAGGT GAGGAATCTT 780

AATAACATCT AATGTACCTA ATGATTTCAA AAGGTGGGGC CGAGTCTGTA TGAAAATACA 840

GATTGTCGCA TTATTTTCTG ATAAGGAGGT GCATTAGAAA CAAGTTAAGC GCACAACTGT 900

GATTACTTAA TTACGCTTTG TGTTCAAGCA GTGAAAGTAA TATGTGAATT GCTTTATCAT 960

GGGAAGATTC AGATATATAT ACATATTCAG TTGTGGCTCA AGGTTTTCTG TAGATATTAT 1020

AGTATTTCAA ATCCATTTCT TCAATATTCC GGATTAGCTC AACACACCCA TTTTTACATG 1080

TTTATAGCCC TTTTTGTGCG CCGTCTTTAA GCCGAAGTGT GAAAAAATCA ACGGCGAGGA 1140

CATGGTCTAC CTGCCATCTT ACGAGATGTG CCGAATTACC ATGGAACCCT GTCGCATTTT 1200

GTACAACACG ACGTTTTTCC CAAAATTCCT TCGCTGCAAC GAAACACTCT TTCCGACGAA 1260

ATGCACAAAC GGAGCACGAG GAATGAAATT CAACGGAACT GGCCAGTGTT TAAGTCCTTT 1320

GGTTCCGACA GATACGTCAG CCAGCTATTA TCCTGGCATC GAGGGCTGCG GCGTGCGATG 1380

CAAGGATCCA CTCTATACCG ATGATGAGCA TCGGCAGATC CACAAACTGA TCGGATGGGC 1440

TGGCAGCATA TGTCTTCTGT CTAACCTTTT CGTGGTGTCC ACCTTCTTCA TCGACTGGAA 1500

GAATGCCAAC AAGTATCCGG CAGTAATTGT GTTCTACATA AATCTTTGCT TTCTAATTGC 1560

TTGCGTCGGG TAAGTTTTGA GCACTATTTT GCATTTGTAT TCTTAATCAA CCACGTATTT 1620

ACTATGCAGC TGGTTGCTTC AGTTTACTTC TGGCTCGCGA GAGGACATAG TATGTCGTAA 1680

GGATGGAACA CTTCGCCACT CAGAGCCTAC AGCCGGTGAA AATCTTTCTT GCATAGTGAT 1740

CTTTGTGCTG GTCTATTATT TTCTCACCGC TGGAATGGTT TGGTTTGTGT TCCTCACCTA 1800

CGCCTGGCAT TGGAGGGCCA TGGGCCACGT CCAAGATCGG ATAGATAAGA AAGGTTCCTA 1860

CTTTCACCTC GTGGCGTGGT CACTACCCCT TGTGCTTACC ATTACCACGA TGGCTTTCAG 1920

TGAGGTGGAT GGAAATAGTA TTGTGGGCAT CTGCTTCGTA GGCTATATCA ATCATTCTAT 1980

GAGGGCAGGA CTACTTCTTG GTCCGCTCTC CGGGGTCATC CTCATTGGTG GATACTTCAT 2040

CACCCGCGGC ATGGTGATGC TTTTTGGACT GAAACACTTC GCTAATGACA TTAAATCAAC 2100

TTCGGCGAGC AACAAAATCC ATTTGATCAT CATGCGCATG GGAGTCTGTG CTCTGCTCAC 2160

TTTAGTTTTC ATACTAGTGG CCATTGCGTG CCACGTTACG GAGTTTAGGC ATGCAGACGA 2220

ATGGGCCCAG AGCTTCAGAC AGTTTATAAT GTAAGTGTAA ATACTCGTTT ATAACTTTTT 2280

CCATACGCCT CAACTAGATA ATTCTTTTGT TTAGCTGCAA AATTTCTTCA GTTTTTGAAG 2340

AAAAGAGTTC CTGTCGAATT GAAAACCGAC CTAGTGTTGG CGTTCTTCAA TTGCATTTGC 2400

TGTGTCTATT TAGCTCTGGA ATCGTAATGT CCACCTGGTG CTGGACACCT TCTTCAATTG 2460

AGACTTGGAA GCGTTATATA AGGAAGTATG TTTTACCATC CTACCAAAAG TTGCATTAAA 2520

GTTATCTGTA CTATCGATTT TATAACCTTT GCAGAAAGTG TGGCAAAGAG GTGGTCGAAG 2580

AAGTGAAAAT GCCGAAGCAC AAGGTCATTG CCCAGACATG GGCCAAGCGC AAGGATTTCG 2640

AGGACAAGGG CAGGCTCTCC ATAACGCTCT ACAACACCCA CACAGATCCC GTGGGGCTCA 2700

ACTTCGATGT GAACGATCTG AACTCTTCTG AGACGAATGA CATCTCATCA ACTTGGGCTG 2760

CATACCTCCC GCAGTGCGTA AAACGTCGCA TGGCTTTGAC GGGAGCAGCG ACAGGTAACT 2820

CGTCAAGCCA TGGACCGCGA AAAAATTCAT TGGATTCCGA GATAAGTGTG AGTGTTCGAC 2880

ATGTTTCCGT TGAATCCCGC AGAAATTCGG TGGACTCGCA GGTATCAGTG AAAATAGCTG 2940

AAATGAAGAC CAAAGTGGCG TCCAGATCAA GGGGAAAACA CGGAGGCTCT TCCAGCAACA 3000

GAAGAACCCA AAGGAGAAGG GATTATATAG CAGCTGCCAC TGGAAAAAGC AGTAGGAGAA 3060

GGGAAAGCAG TACTTCAGTG GAGTCGCAGG TCATCGCGCT CAAGAAAACG ACCTATCCCA 3120

ATGCTAGTCA CAAAGTGGGC GTGTTTGCTC ATCACAGCTC CAAGAAACAA CACAATTACA 3180

CCAGCTCCAT GAAGCGAAGG ACTGCTAATG CCGGATTGGA TCCCTCTATT CTTAATGAAT 3240

TCCTGCAGAA AAATGGCGAT TTTATATTCC CATTCCTCCA AAATCAAGAT ATGAGCTCTA 3300

GTTCGGAGGA GGATAATTCC AGAGCATCCC AAAAGATTCA GGATCTTAAC GTGGTTGTAA 3360

AGCAGCAGGA AATAAGTGAG GATGATCACG ACGGAATAAA GATTGAAGAA CTGCCAAATA 3420

GCAAACAGGT GGCATTGGAG AACTTTCTTA AAAACATAAA AAAATCTAAT GAATCCAATT 3480

CTAACCGACA TTCCCGAAAT TCCGCAAGAA GTCAGTCAAA AAAGTCCCAA AAGAGACATC 3540

TCAAGAACCC TGCTGCTGAT CTAGATTTCA GGAAGGACTG TGTAAAGTAT CGGTCTAATG 3600

ACTCACTTAG CTGCTCCTCT GAAGAGCTGG ATGTGGCTTT GGACGTAGGA AGCCTTCTTA 3660

ACAGCTCTTT TTCTGGAATA TCCATGGGCA AACCACATAG TAGAAACAGC AAAACCAGCT 3720

GCGATGTGGG CATACAGGCT AATCCTTTCG AGCTAGTTCC CAGTTACGGA GAAGACGAAC 3780

TGCAGCAGGC CATGCGACTC CTAAACGCAG CCAGCAGACA AAGAACTGAA GCAGCCAATG 3840

AGGATTTCGG AGGAACGGAG CTGCAGGGCT TGTTGGGTCA TTCCCATCGG CATCAAAGGG 3900

AGCCCACGTT TATGAGCGAG TCGGACAAAC TCAAAATGTT ATTGCTGCCT TCAAAATAGC 3960

AAGACTAAAT AAGCAATTGA TGCATTTACT TAAGGTTCAA AAACTCTTAC AATATTGTAG 4020

TTTTTGTTCT AAGAAATCAA ATTGTTAGCG CTGAAAATAA TCGTACAATC TTATCTATTT 4080

TACGAAATCG TAATATTGTT ATGTTCACTG TTCAACGATT TATAAGAATA TATCGCTTCA 4140

CTAGAATTGG AAACCCAAAT GATATTTAAA ACAAACAAAT ACGAAATTGT AGTACACAAG 4200

CCAGAGCAGT TTACATGCGA TGAACATTTA GATTCTTCTT AATCGATTAC TGGAACAGAC 4260

TGAGCGAAAC TAGAACTACG AATTACGAAT ACTCATAGTC ATTAGGCTGC AACTTTATTT 4320

TACAGATTCA TCACCCCATC TAGCTTGTAA GCATTCGAAT CTCTGTGTAC GTTTGTGAAT 4380

GACTGTTTCC TTAATCCTGG TACTCACGCC AAAGTAAATG CCAAAGAGGA TAATAATTTA 4440

TTTTCATTAT TTTTCTTTGC CGTGGGTACA GGACTTTAGA TTGTAGATTA TAGATTTAAG 4500

TACGATATAA ATAAGCTTCT TGGGCACACA AATCGTACCT CAGAAAGTGC CTTCAAGTTT 4560

ACAAAATTAT ACATAATAAT TTGTGTAACT AATAAACGAT TTTAAATCCT CGAGTCT 4617